AUGUST
1961
VOLUME 25
NUMBER 4

JOURNAL OF CONSULTING PSYCHOLOGY

AMERICAN PSYCHOLOGICAL ASSOCIATION

JOURNAL OF CONSULTING PSYCHOLOGY

EDWARD S. BORDIN, Editor
University of Michigan

BOYD MCCANDLESS, Associate Editor
University of Iowa

Advisory Editors: John W. Atkinson, *University of Michigan*. Frank X. Barron, *University of California, Berkeley*. John M. Butler, *University of Chicago*. Allen T. Dittmann, *National Institute of Mental Health*. Seymour Fisher, *Baylor University*. Harrison G. Gough, *University of California, Berkeley*. William L. Hays, *University of Michigan*. Robert B. Malmo, *McGill University*. Paul E. Meehl, *University of Minnesota*. Daniel R. Miller, *University of Michigan*. Alan K. Rosenwald, *University of Illinois*. Saul Rosenzweig, *Washington University*. Julian B. Rotter, *Ohio State University*. Nevitt Sanford, *University of California, Berkeley*. Seymour B. Sarason, *Yale University*. Robert R. Sears, *Stanford University*. Edwin S. Shneidman, *Veterans Administration, Los Angeles, California*. Read D. Tuddenham, *University of California, Berkeley*

The *Journal of Consulting Psychology* is the clinical journal of the American Psychological Association. It is devoted primarily to the publication of original research relevant to psychological diagnosis, psychotherapy and counseling, personality, and the dynamics of behavior. Case studies, theoretical contributions, descriptions of clinical techniques, and discussions of the professional problems of clinical psychology also appear in its pages.

Manuscripts submitted for *regular* or *early* publication and manuscripts for Brief Reports should be sent to the Editor: Edward S. Bordin, Department of Psychology, University of Michigan; Ann Arbor, Michigan. All manuscripts should be submitted *in duplicate* to facilitate referral to the Advisory Editors.

Style of Manuscripts. Manuscripts must be prepared in the style described in the *Publication Manual of the American Psychological Association (1957 Revision)*, obtainable for $1.00 from the Washington office of the APA.

Early Publication. Authors may secure the early publication of an acceptable article in the first issue to go to press after its acceptance by arranging for the article to be printed as extra pages, increasing the number of pages that subscribers receive. Information about charges for early publication may be obtained from the Managing Editor.

American Documentation Institute. To save printing costs, materials supplementary to articles may be deposited with the ADI. To obtain copies, order by Document Number as cited in the reference footnote, from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. The indicated fee must be remitted in advance. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Reprints. Authors receive fifty reprints gratis. Additional quantities may be ordered when returning proofs.

Arthur C. Hoffman
Managing Editor

Helen Orr
Promotion Manager

Communications concerning subscriptions, change of address, claims for the nonreceipt of a number, advertising, and other business matters, should be sent to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D.C. Address changes must reach the subscription office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

The *Journal of Consulting Psychology* is published bimonthly in February, April, June, August, October, and December, the yearly volume comprising approximately 576 pages. The subscription per year is $10.00, foreign $10.50, single number $2.00.

Published by the American Psychological Association, Inc.
at Prince and Lemon Sts., Lancaster, Pa.
and 1333 Sixteenth St. N.W., Washington, D.C.

Second-class postage paid at Lancaster, Pa.
© 1961 by the American Psychological Association, Inc.

JOURNAL OF CONSULTING PSYCHOLOGY

August 1961 Vol. 25, No. 4

Contents

ARTICLES

A Study of Reading Disability: *Richard H. Walters, Malle Van Loan, and Irene Crofts* 277

Psychologists' Judgments of Physical Handicap from H-T-P Drawings: *Orval G. Johnson and Frank Wawrzaszek* 284

Changes in Intellectual Functions of Children in a Psychiatric Hospital: *E. Wesley Hiler and David Nesvig* 288

A Comparison of Social and Solitary Male Delinquents: *Mary H. Randolph, Harold Richardson, and Ronald C. Johnson* 293

The Dimensionality of Ratings of Therapist Verbal Responses: *Edmund S. Howe and Benjamin Pope* 296

Need Value and Expectancy Interrelations as Assessed from Motivational Patterns of Parents and Their Children: *Forrest B. Tyler, Bonnie B. Tyler, and Janet E. Rafferty* 304

Cross-Validation of a Rorschach Checklist Associated with Suicidal Tendencies: *Irving B. Weiner* 312

Concurrent and Construct Validity of Direct and Indirect Measures of Dependency: *Marvin Zuckerman, Eugene E. Levitt, and Bernard Lubin* 316

Response Bias in Questionnaire Reports: *N. H. Azrin, W. Holz, and I. Goldiamond* 324

The MMPI as a Measure of Character Structure as Revealed by Factor Analysis: *Joseph C. Finney* 327

Play Therapy Limits and Theoretical Orientation: *Haim G. Ginott and Dell Lebo* 337

The Construct Validity of the Edwards PPS Heterosexuality Scale: *E. Jerry Phares and Calvin K. Adams* 341

Second-Order Personality Factor Structure in the Objective Test Realm: *R. B. Cattell, R. R. Knapp, and I. H. Scheier* 345

Level of Aspiration in Hypertensive Cardiac Patients Compared with Nonhypertensive Cardiac Patients with Arteriosclerotic Heart Disease: *Desmond D. O'Connell and Richard M. Lundy* 353

Objective Estimates of Clinical Judgments: *Rolfe LaForge* 360

A Note on "Impulse Repression and Emotional Adjustment": *H. J. Eysenck* 362

Sex Differences in Mental Health Analysis Scores of Elementary Pupils: *Joseph C. Bledsoe* 364

BRIEF REPORTS

A Comparison of Acceptors and Resistors of Drug Treatment as an Adjunct to Psychotherapy: *Allen Raskin* 366

An Interpretation of *m: John M. Reisman* 367

The Relation of the Famous Sayings Test to Self- and Ideal-Self-Adjustment: *Bernard I. Murstein* 368

The Intelligence of Boys with Muscular Dystrophy: *Don Keith Worden* 369

Symbolic Interpretation of Rorschach Content: *Joseph F. Rychlak and Donald E. Guinouard* 370

Brief Reports

The *Journal of Consulting Psychology* will accept Brief Reports of research studies in clinical psychology for early publication without expense to the author. The procedure is intended to permit the publication of soundly designed studies of specialized interest or limited importance which cannot now be accepted because of lack of space. Several pages in each issue will be devoted to Brief Reports, published in the order of their receipt without respect to the dates of receipt of the regular articles. Most Brief Reports appear in the first or second issue to go to press following their final acceptance.

An author who wishes to submit a Brief Report:

1. Sends the Brief Report, limited to one printed page and prepared according to the specification given below.

2. Also sends to the Editor a full report of the research study, in sufficient detail to give a clear account of its background, procedure, results, and conclusions, which will be filed with the American Documentation Institute to insure indefinite availability.

3. Prepares at least 100 mimeographed copies of the full report, which the author will send without charge to all who request it as long as the supply lasts.

4. Agrees not to submit the full report to another journal of general circulation.

Specifications

Brief Report. The Brief Report should give a clear, condensed summary of the procedure of the study and as full an account of the results as space permits.

To insure that the Brief Report will be no longer than one printed page, its typescript, including all matter except the title and the author's lines, must not exceed 85 lines averaging 42 characters and spaces in length. Set the typewriter margins for short lines of 42 characters, which are 3.5 inches long in elite typing, and 4.2 inches long in pica.

The manuscript of the Brief Report must be *double spaced* throughout. Except for its short lines, it follows the standard style of the 1957 revision of the *APA Publication Manual.* Headings, tables, and references are avoided or, if essential, must be counted in the 85 lines. Each Brief Report must be accompanied by a footnote in the style below, which is typed on a separate sheet and *not* counted in the 85-line quota:

[1] An extended report of this study may be obtained without charge from John Doe (giving the author's full name and address) or for a fee from the American Documentation Institute. Order Document No. —— from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $—— for microfilm or $—— for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

Extended report. Because the extended report is intended for photoduplication, and is not copy to be sent to a printer, its style should differ in several ways from that of other manuscripts: (*a*) The extended report should be typed with single spacing for economy in duplication. (*b*) Tables and figures should be placed adjacent to the text which refers to them. A caption should be typed below each figure. (*c*) Footnotes should be typed at the bottom of the page on which reference is made to them. In other respects, the full report is prepared in the style specified by the *Publication Manual.*

Journal of Consulting Psychology
1961, Vol. 25, No. 4, 277–283

A STUDY OF READING DISABILITY

RICHARD H. WALTERS,[1] MALLE VAN LOAN, AND IRENE CROFTS[2]
University of Toronto

The psychoanalytic theory of reading disability (Blanchard, 1946; Fenichel, 1937; Strachey, 1930), which is closely related to Freud's (1924) theory concerning hysterical blindness, has received relatively little attention from research psychologists in spite of its popularity among psychoanalytically oriented therapists. In a recent outline of this theory, Jarvis (1958) has drawn attention to three factors that supposedly characterize the retarded reader: an avoidance of looking, the problem of aggression, and faulty identification mechanisms. Jarvis regards the "active part of looking" as creating the major difficulty for the retarded reader; the counterpart of this activity outside the school "is an inability to identify predominantly with one's own sex" (p. 468).

The studies reported in this paper were suggested by the psychoanalytic theory of reading disability. Since it is the alleged sexual significance of reading that, according to psychoanalysts, gives rise to fear or avoidance of looking, it was hypothesized that retarded readers would show greater hesitation in looking at a sexual object than would either average or advanced readers, and that, through generalization, this "avoidance" would also be evident in perceptual tasks involving identification of nonsexual objects. It was further hypothesized that retarded readers would display hostility toward the same-sex parent, a condition that might reflect both aggression and "faulty identification mechanisms."[3]

[1] The authors wish to express their appreciation to the Scarborough Board of Education and to the staffs of the R. H. King Collegiate and Corvette Schools, especially to Mathilde Ziehr and James Phillips, for their co-operation in the studies reported in this paper. They are also grateful to Marian B. Hyatt for her assistance in testing.

[2] Now at the Child Guidance Clinic of Greater Winnipeg.

[3] A preliminary study of the relationship between reading achievement and identification produced promising results for boys, but negative results for girls. Mimeographed copies of this study may be obtained from the senior author.

METHOD

Subjects

Ss were 58 Grade 3–6 boys from a single suburban school. They were selected from a larger group of 86 boys who were of average intelligence (IQs between 88 and 112) and who were, according to their medical histories, free of eye anomalies, hearing losses, and behavior problems. Of the 86 boys, 20 had reading ages that were at least 1 year below their mental ages; these boys were regarded as retarded readers. Eighteen boys whose reading ages were at least 1 year above their mental ages were regarded as advanced readers. From the remaining 48 boys, the 20 boys with the smallest discrepancies (less than 6 months) between reading age and mental age were selected as average readers.

Reading ages and mental ages were available in the school records; since the reading tests and intelligence tests had, in some cases, been administered several months apart, the mental ages were corrected to correspond to the time at which the reading test was given. Unfortunately, the tests administered varied to some extent from grade to grade. However, since the administration of further tests would have too greatly disrupted the school timetable, the available indices were accepted as reasonably adequate bases for group selection. The tests given were, in fact, very similar, e.g., the Otis group test and the Dominion Test of Learning Capacity as measures of intelligence and the Gates Vocabulary, Shenell Vocabulary, and Dominion Silent Reading Tests as measures of reading ability.

Apparatus and Tests

Two perceptual tests[4] developed for the Cerebral Palsy Project of the Department of Psychology, University of Toronto, were included in the test battery. The Steer-Allen Figure-Ground Confusion Test consists of 15 cards, on each of which is depicted an object whose outline is interrupted by the background. In order to identify the object, *S* must free the figure from the confusing background detail. Since this test was developed for studies of children younger than those used in this study, it was employed primarily as a rapport building test at the commencement of

[4] The authors are indebted to H. O. Steer both for permission to use these tests and for his helpful advice on this study.

FIG. 1. Multiple-choice apparatus, showing nude doll used as test object.

testing. The Steer-Beatty Closure-Threshold Test, a more suitable test for children in Grades 3–6, consists of 12 sets of five cards, each card being composed of dots. Each set forms a graduated series of representations of the same object. The degree of clarity of the representations is a function of the relative spacing of dots that compose the figure and dots that compose the background. The cards in each set are presented consecutively in an order that makes the identification of the object (figure) increasingly less difficult. These perceptual tests were included because they appeared to require "active looking" in the sense in which this term is used by psychoanalysts.

A multiple-choice apparatus (Figure 1) was used for the major test of the "fear of looking" hypothesis. A box, approximately 3 feet long and 1½ feet high, was separated into four compartments. Each compartment was fronted by a separate door which opened easily. The back of the box also opened to allow the experimenter (*E*) to insert an object into any one of the compartments. The doors and the back of the box were connected with a buzzer and an electric timer. When the box was completely closed, back and front, the electrical circuit was completed and both the buzzer and timer were set in operation. Opening any one of the doors interrupted the circuit and shut off both the buzzer and the timing mechanism. This apparatus provided an automatic recording of the interval between *E*'s signal for *S* to respond (buzzer) and *S*'s opening of one of the doors to look at a hidden object. Three objects were used during testing: a female Dutch doll of a conventional type (neutral object), a nude male doll with a penis (the type sometimes employed by psychoanalytically oriented child psychiatrists), and clothed boy doll (a second neutral object).[5]

Since psychoanalysts have stressed the importance of "underlying" unconscious psychological determinants, and some researchers (e.g., Friedman, 1952) have claimed that these can be diagnosed only through the use of projective techniques, a brief picture-story test was added to the test battery. Four pictures were included: [6]

1. A boy is shown turning his back on an older male figure, who is walking away in the opposite direction.

2. A boy is shown looking into a bathroom. An arm of a taller figure protrudes from the shower curtain, and water can be seen coming from the rose of the shower.

3. A boy stands in front of an older seated male figure, who is reading a newspaper. The boy is gesticulating with arms outstretched.

4. A boy is shown looking into a room in which an adult male and female are embracing.

Pictures 1 and 3 were chosen to test hostility toward the father; Pictures 2 and 4 to test fear and avoidance of looking.

An attempt was made to assess identification by a modified version of Fiedler's (1958) Assumed Similarity to Others (ASo) Test. This test, however, proved to be beyond the comprehension of the younger children in the study and, consequently, the results were of little value. A simple test of parent preference, described below, was also given.

Procedure

S was brought to the experimental room by a female *E* and was seated at a desk facing the discrimination apparatus. *E* seated herself to the rear of the apparatus about 4 feet away from *S*.

E first presented the Figure-Ground Confusion Test, using the following instructions: "I am going to show you some pictures in each of which a thing is hidden. I want you to tell me what that hidden thing is." *S* was given a trial run with a card representing a bird, after which *E* pointed out the detail of the bird. The remaining pictures were then presented one at a time. No time limit was set. If *S* said that he could not find the object, *E* simply presented the next card. *S*'s responses were recorded on data sheets.

S was now asked to stand in front of the discrimination apparatus and was given the following instructions:

> Here I have this doll [*E* held up the Dutch doll] and I am going to hide it in one of these boxes. I want you to see if you can find it. You can do this by opening one of these doors [demonstrated] as soon as you hear the buzzer. When you hear the buzzer you can open the door in which you think I have hidden the doll. Don't close the door again until I tell you to.

S was given three trial runs with the Dutch doll, then the experiment proper was begun. The doll was placed into the various compartments from the back of the apparatus in a predetermined random sequence. *S* was given 10 trials with this doll. The

[5] The choice of a nude male doll as the test object was based on psychoanalytic symbolism (Fenichel, 1937; Freud, 1953; Strachey, 1930) concerning the act of looking and the mastery of reading. The play therapy dolls were lent by Alice Moultby of the York Township Child and Adolescent Guidance Clinic.

[6] These pictures form part of a series developed by Albert Bandura of Stanford University.

latency and choice of box in each trial were recorded by *E*.

E now took out the nude father doll, holding it up in the air in full view of *S*. *E* said: "Look at this doll! Now I am going to do the same thing with this doll, and I want you to do exactly as you did before." Again the doll was placed, in a predetermined random sequence, into each of the compartments. Ten trials were given with the father doll, and the results were recorded as before.

Finally, *E* took the clothed boy doll. Holding it up, *E* said: "Now I am going to use this doll, and I want you to do exactly the same as before." This time the doll was placed in each compartment in an orderly sequence: 1–2–3–4, 1–2–3–4. The purpose of this final set of trials was twofold. It was thought that the subsequent presentation of a neutral object might reduce any emotional upset produced by the nude doll; in addition, through the use of a regular sequence, it seemed likely that *S* would finish up by making some "correct" responses, so reducing possible feelings of failure. The plan was to continue hiding the boy doll until *S* had made two successive correct responses. Within the time limit imposed by the school schedule, this was not possible in all cases.

The design of the experiment would have been improved if presentations of a neutral doll and the nude doll had been made in random order. However, to minimize the possibility of emotional upset and of subsequent parental complaints about the use of a nude figure, it was thought wiser to buttress the presentations of the nude figure with preceding and subsequent presentations of neutral figures.

S was now taken to a second female *E*, who had not been associated with the presentation of the nude figure. *E* seated *S* beside her and took out a quarter. She said:

> Now I want you to imagine that you are going on a long, long trip. You can go with your mother or your father, but you cannot go with both. I am going to toss this coin. "Heads" you go with your father, "tails" you go with your mother. You call.

A preliminary trial was given to insure that *S* complied with instructions. The instructions were then carefully repeated, and a further trial was given. *S*'s response was recorded on each trial.

E now gave the picture-story test. *S* was instructed:

> I want you to make me up a story about this picture. Tell me what the little boy is doing, what led up to this, how the little boy is thinking and feeling, and how it will all turn out.

The only probes given were repetitions of parts of the original request or variations on these. *S*'s stories were recorded on tape and later transcribed.

The abortive ASo test was then administered. Finally, *E* asked the *S* to stand at a point in the room approximately 8 feet away from her and presented the Closure-Threshold Test. The sets of cards were presented in a standard order. *S* was shown the first (most difficult) card in a set and was asked: "What does this look like to you?" If *S* failed on this card, he was shown the next card in the series and was asked what it looked like. The procedure was continued until *S* responded correctly or until all five cards in a series had been presented. On Cards II to V, if *S* said he did not know what the card looked like or that it did not look like anything, *E* said: "Perhaps you can tell me what it looks *most* like. What do you think it might be?" The procedure was continued until all 12 sets of cards had been shown.

The testing was now complete, and *S* was taken back to his classroom.

TABLE 1

DISTRIBUTION OF ERROR-FREE RECORDS ON THE FIGURE-GROUND CONFUSION TEST

Errors	Advanced Readers	Average Readers	Retarded Readers
Present	7	7	2
Absent	11	13	18

Note.—Chi square = 4.797; $p < .10$.

RESULTS

Avoidance of Looking

As expected, a large number of *S*s obtained a perfectly correct record on the Figure-Ground Confusion Test. Consequently, a chi square analysis was performed with the data divided into two categories—errors present and errors absent. The distribution of cases among the three groups of *S*s is given in Table 1. The results were in the predicted direction; however, the distribution could have occurred by chance 10 times in 100.

Results of the Closure-Threshold Test were assessed as follows. For each set of cards, *S*'s score consisted of the number of cards required to elicit a correct response; if *S* failed

TABLE 2

ANALYSIS OF VARIANCE BY RANKS OF RESPONSES TO THE CLOSURE-THRESHOLD TEST

Sums of Ranks			
Advanced Readers ($N = 18$)	Average Readers ($N = 20$)	Retarded Readers ($N = 20$)	H
429.5	580.0	701.5	6.219*

* $p < .05$.

TABLE 3

ADJUSTED GROUP MEAN LATENCIES (IN SECONDS) OF RESPONSES TO THE NUDE DOLL

Advanced Readers ($N = 18$)	Average Readers ($N = 20$)	Retarded Readers ($N = 20$)
1.417	1.744	2.044

TABLE 4

ANALYSIS OF COVARIANCE OF LATENCIES OF RESPONSES TO NUDE DOLL

Source	Adjusted *SS*	*df*	Adjusted *MS*	*F*
Between groups	3.137	2	1.568	10.181**
Within groups	8.331	54	0.154	

** $p < .001$.

to respond correctly to the fifth (the easiest) card, he was arbitrarily assigned a score of 6. These component scores were summed to provide a total score for each *S*. Total scores were then ranked for a Kruskall-Wallis analysis of variance by ranks (Siegel, 1956). Results were as predicted (Table 2).

The median latency of response to the first two dolls in the multiple-choice task was computed for each *S*. Medians were preferred to means as a measure of central tendency because of the possible presence of a single deviant response within a series of trials. Using the responses to the nude doll as the dependent variable, an analysis of covariance was carried out with responses to the Dutch doll as the covariant. Table 3 gives the adjusted group mean latencies to the nude doll, and Table 4 the results of the analysis of covariance. A test of homogeneity of regression had previously indicated that this procedure was justifiable ($F = 1.95$; $p > .05$). Differences were in the predicted direction and were significant at the .001 level. Subsequent t tests indicated that this result was largely due to the superior performance of the advanced readers, whose performance was significantly better ($p < .001$, one-tailed test) than that of either of the other two groups. Because all three groups of *S*s tended to show some decrease in latency in response to the nude doll (undoubtedly a practice effect), a further analysis of covariance of responses to the nude doll was carried out, this time with responses to the boy doll as the covariant. Significant differences in the predicted direction were obtained ($F = 6.46$; $p < .005$).

Parent Preference Test

The results of the coin tossing test (second trial) are given in Table 5. The percentage of retarded readers showing preference for the same-sex parent was smaller than that of either of the other two groups ($p < .02$). When results from the preliminary trial were included in the analysis, the differences among the three groups were considerably reduced (Table 6).

Picture-Story Test

The stories were scored from typescripts by an assistant who did not know to which group individual *S*s belonged. Instructions for scoring were as follows:

Story 1. Score + if taller figure is identified as the father and if the boy in the picture expresses hostility toward the father (i.e., is described as "angry," "mad," or as wishing to harm the father).

TABLE 5

DISTRIBUTION OF PREFERENCES ON COIN-TOSSING TEST: SECOND TRIAL

Preference	Advanced Readers	Average Readers	Retarded Readers
For father	13	10	4
For mother	5	10	16

Note.—Chi square = 10.561; $p < .01$.

TABLE 6

DISTRIBUTION OF PREFERENCES ON COIN-TOSSING TEST: COMBINED TRIALS

Preference	Advanced Readers	Average Readers	Retarded Readers
For father on both trials	9	7	3
For mother on one or both trials	9	13	17

Note.—Chi square = 5.339; $p < .10$.

Story 2. Score + if the boy in the picture (*a*) does not mention the figure behind the shower curtain, or (*b*) is described as unwilling to enter the bathroom because someone is already in there, or (*c*) is described as being afraid, ashamed, or guilty, because he has seen the figure in the shower.

Story 3. Score + if seated figure is identified as the father and if the boy expresses hostility toward the father (as for Story 1).

Story 4. Score + if the boy in the picture (*a*) does not describe the man and woman as kissing or making love, or (*b*) is described as being afraid, ashamed, or guilty because he has seen the adult figure making love.

For the purpose of analysis, *Ss* were regarded as showing hostility toward the father if they received a + score on either Story 1 or Story 3, and as showing fear of looking if they received a + score on either Story 2 or Story 4. Subsequent chi square tests failed to support the hypotheses being tested.

Discussion

Only the "fear of looking" tests could provide crucial support for the psychoanalytic theory. One of these, the picture-story test, yielded completely negative results. The multiple-choice task, on the other hand, may be interpreted as supporting the psychoanalytic theory. However, in view of the negative results of the picture-story test, alternative interpretations must be favored.

Since the differences between advanced readers and both the other groups of *Ss* on the multiple-choice task were significant at the .001 level, interpretation could depend heavily upon the characteristics of over-achievers, and particularly upon information about child training practices that tend to produce high-achieving children. Unfortunately, little is known about the family backgrounds of children who are exceptionally advanced in reading. On the other hand, the studies of McClelland, Atkinson, Clark, and Lowell (1953) do give some information concerning the family backgrounds of *Ss* with high need achievement. It is possible that over-achieving readers fall within the larger group of individuals who are highly oriented toward achievement in general. Making this assumption, one might expect that over-achieving readers come from homes in which there is considerable stress on independence training, in which the parents are democratic rather than autocratic, and in which conventional moral standards are not highly stressed. It is this latter factor that may be important in interpreting the results of this experiment. A child who is not deterred from peering into a compartment that, if his choice is correct, contains a nude figure with genitals may thus be the product of a home in which conventional moral standards are not stressed.

Observance of conventional moral standards in the area of sex training involves considerable emphasis on modesty (Sears, Maccoby, & Levin, 1957). If the emphasis on modesty training is not so strong in the homes of high achieving children, one might expect that these children have had opportunity to see their same-sex siblings and fathers in the nude. Therefore one would not expect the over-achieving children in this study to be greatly deterred from responding quickly in a task involving a nude doll of the same sex as themselves.

McClelland et al. (1957) have found that individuals with high need achievement tend to come from homes in which the parents make strong demands for independence, including strong achievement demands. Thus, these children may not only be relatively uninhibited about sexual matters, but may also have been reinforced for exploratory behavior. This latter factor, along with low emphasis on modesty training, may be partly responsible for the obtained difference between the over-achievers and the other two groups of *Ss*.

Since, for all three groups of *Ss*, there was a decrease in median latency over the three sets of trials, it seemed possible that the inclusion of the nude doll as a test object was, in fact, of little importance, and that the results of the discrimination test merely reflected differences in learning capacity in a perceptual-motor task. This interpretation is partly borne out by Figure 2. In this figure changes in test object (Dutch doll, nude doll, boy doll) are ignored. The median response of each *S* on each block of 5 trials was first identified. These median latencies were then averaged to provide an indication of changes in latency over a series of 30 trials. The figure indicates that advanced readers improve

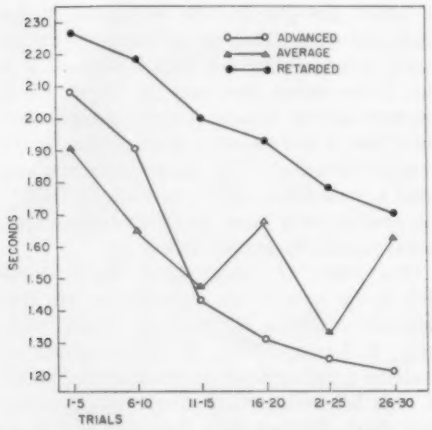

FIG. 2. Change in latencies of advanced, average, and retarded readers over a series of 30 trials, for all stimuli in the "looking" test.

in performance at a much faster rate than do retarded readers. Average readers show an initial improvement that is intermediate between that of the other two groups, then perform somewhat erratically. These results suggest the necessity for a further test in which a single test object is utilized and in which trials are continued until asymptotes are reached for all three groups.

On the two perceptual tests the retarded readers performed more poorly than both the average and the advanced readers. Once again, these results could be interpreted as supporting the psychoanalytic theory. A simpler explanation, however, is that perceptual skills are highly developed among advanced readers and poorly developed among retarded readers, and that the development of these skills is related to reinforcement of exploratory behavior by care-taking adults.

The above explanation of the findings imply that fear of looking is not a *causative* factor producing differences among the three groups of *Ss* in their responses to the nude doll. It is suggested that, at the most, parents who inhibit exploratory behavior are also nonpermissive in their modesty training and that, as a consequence, children who are retarded readers tend also to be sexually inhibited. In this connection it is important to remember that during a child's early years sexual behavior largely occurs in the form of curiosity or exploratory behavior involving perceptual-motor responses.

In an attempt to integrate the findings concerning hostility and identification into a tentative theory concerning the antecedents of reading disability, the fantasy data will be ignored on the grounds that, in spite of their widespread utilization in clinical settings, such data seem to have no consistent relationship with supposedly corresponding overt responses. The coin tossing test suggests that boys who are retarded readers are relatively hostile toward their fathers.

Let us now make the further assumption, for which child training studies afford some justification, that mothers tend to be somewhat nonpermissive concerning exploratory behavior and that fathers are more variable in this respect. In this case, the amount of reinforcement which exploratory behavior receives will depend considerably upon the father's behavior patterns. Hostility to the father may then be viewed as an outgrowth of paternal nonpermissiveness and punitiveness, i.e., frustration, of exploratory behavior, of which sexual behavior is an important facet.

SUMMARY

Psychoanalysts have attributed reading disability to three, supposedly related, factors: fear and avoidance of looking; hostility, primarily toward the same-sex parent; and failure to identify with the same-sex parent.

Hypotheses suggested by psychoanalytic theory were investigated in a study in which Grade 3–6 boys were employed as *Ss*. Retarded readers performed more poorly on two perceptual tasks involving recognition of form than did average and advanced readers. They were slower in opening a compartment to look for a male nude doll than were the other two groups. In addition, they chose their father less often on a simple parent preference test. On the other hand, fantasy data failed to support either the "fear of looking" or the hostility hypothesis.

Some of the above results may be interpreted as supporting the psychoanalytic theory. An alternative interpretation in terms of parental conditioning of exploratory and sexual responses was nevertheless favored.

REFERENCES

BLANCHARD, PHYLLIS. Psychoanalytic contributions to the problem of reading disability. *Psychoanal. Study Child.*, 1946, **2**, 163–168.

FENICHEL, O. The scopophilic instinct and identification. *Int. J. Psycho-Anal.*, 1937, **18**, 6–34.

FIEDLER, F. E. *Leader attitude and group effectiveness.* Urbana, Ill.: Univer. Illinois Press, 1958.

FREUD, S. Psychogenic visual disturbance according to psychoanalytic concepts. In E. Jones (Ed.), *Collected papers of Sigmund Freud.* Vol. 2. London: Hogarth, 1924.

FREUD, S. The interpretation of dreams. In J. Strachey (Ed.), *The standard edition of the complete psychological works of Sigmund Freud.* London: Hogarth, 1953.

FRIEDMAN, S. M. An empirical study of the castration and oedipus complexes. *Genet. psychol. Mongr.*, 1952, **46**, 61–130.

JARVIS, V. Clinical observations on the visual problem in reading disability. *Psychoanal. Study Child.*, 1958, **13**, 451–470.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

SEARS, R. R., MACCOBY, ELEANOR E., & LEVIN, H. *Patterns of child rearing.* Evanston, Ill.: Row, Peterson, 1957.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

STRACHEY, J. Some unconscious factors in reading. *Int. J. Psycho-Anal.*, 1930, **11**, 322–331.

(Received June 21, 1960)

Journal of Consulting Psychology
1961, Vol. 25, No. 4, 284–287

PSYCHOLOGISTS' JUDGMENTS OF PHYSICAL HANDICAP FROM H–T–P DRAWINGS

ORVAL G. JOHNSON AND FRANK WAWRZASZEK
Lewis County Schools, Washington *Eastern Michigan University*

The psychological characteristics of the physically handicapped person have been studied increasingly in recent years. Force (1956), Lerner and Martin (1955), Shere (1956), Whitehouse (1953), Wrightstone (1957), and Cruickshank (1955) suggest various degrees and kinds of differences psychologically between handicapped and nonhandicapped persons. On the other hand, Levy and Michelson (1952) and Wenar (1956, 1958) find no differences or only minor differences between the two groups. Berreman (1954) looks at the problem from the standpoint of the attitudes shown toward handicapped people, which he says are different from attitudes toward normals, and are likely to affect the self-images of the handicapped. Wawrzaszek, Johnson, and Sciera (1958) found no differences between handicapped and nonhandicapped children on any of 10 variables derived from the House-Tree-Person Test.

Problem

The purposes of this study are twofold:

1. To determine whether or not physically handicapped children project into their drawings any feelings about their handicaps to the extent that the handicap can be detected by psychologists through an analysis of their drawings

2. To investigate the characteristics of drawings that psychologists use to postdict which of a matched pair of children is handicapped

Method

The H-T-P test was administered to 37 handicapped children in the special classes of a public school. Sixteen were postpolio cases, seven cardiac, seven cerebral palsied (mild), two spina bifida, and one each Perthes hip, muscular dystrophy, congenital deformity, slipped epiphysis, and brittle bones.

These children were selected from a slightly larger group, eliminating those who might be hampered in their drawing by impaired motor coordination. This selection was accomplished with the consultation of the physical therapist.

A control group was made up by matching each of the handicapped on the basis of age, sex, and IQ. The chronological age of each control was within 3 months and the IQ within 10 points of that of the handicapped child. Therefore, by definition, the average chronological age and IQ for both groups were the same. Actual computations showed that there was no significant difference between the two groups in CA and IQ. All but a few of the handicapped and control children had been tested with individual intelligence tests. Table 1 shows the mean, range, and *SD* of CA and IQ for both groups.

The H-T-P test was administered to each child in accordance with the instructions suggested by Buck (1948). The matched pairs of protocols, randomized for order of handicapped and nonhandicapped, were presented to nine psychologists with the following instructions:

> These are H-T-P protocols of 37 pairs of children matched for sex, CA, and IQ. One of the children is physically handicapped and one is his nonhandicapped control. You are to judge from the protocol which is the physically handicapped child and which is the control. Please clip a slip of paper onto the protocol of the handicapped child, noting briefly the basis for your decision in each case.

TABLE 1

Mean, Range, and Standard Deviation of Handicapped and Nonhandicapped Groups on Chronological Age and Intelligence Quotient

Measure	Chronological Age		Intelligence Quotient	
	Handicapped	Nonhandicapped	Handicapped	Nonhandicapped
Mean	116.3	118.3	103.4	104.2
Range	6–5 to 13–6	6–6 to 13–5	63 to 128	63 to 129
SD	25.6	26.1	13.9	14.7

All of the psychologists had experience in working with children, a majority of them having done psychological work with physically handicapped children. The drawings of the Person were evaluated according to the Goodenough (1926) scale, and an MA based thereon derived for each subject. The drawings of the Person were also measured for height.

Results

Of the 333 judgments (9 judges × 37 judgments for each), 208 or 62% were correct. This proportion is significantly different (.01 level) from chance expectation. The percentages correct for the judges individually ranged from 54 to 65. None of these is significantly different from chance.

In many cases there was substantial agreement among the psychologists as to which of the protocols was that of the handicapped child. Sometimes they agreed and were right, and sometimes they agreed and were wrong. If we arbitrarily assume that *more* than two-thirds agreement by the psychologists is substantial agreement, then it can be said that in 27 out of 37 possible cases, or 73%, there was substantial agreement. In 19 out of the 27 cases where substantial agreement occurred, the majority made a correct judgment. Thus, when they agreed they were right in 70% of the cases, wrong in 30%.

Practically all of the reasons given by the judges for classifying a protocol as handicapped or nonhandicapped involved psychodynamic interpretations. Some typical reasons are as follows:

Treatment of end gables of house. Anxiety reflected in shading on trees.

Although I like the person I do feel there is some distortion.

Distorted perspective on house. Anxiety reflected in tree.

Much anxiety in human drawing. Great difficulty in form.

Faltering lines. Hands omitted.

Small person in relation to other figure.

The handling of the branch structure of the tree indicates some confusion; the general quality of the person is inferior.

Insecure house. Height relation to width. Chopped off trunk which implies trauma.

Disability suggested in figure drawing.

To determine whether or not some aspect of the drawings other than the projective characteristics may have influenced the judges, the Goodenough scores were determined by an analysis of the drawing of the Person. Table 2 shows the relation between rightness or wrongness of the majority judgment and the variable of whether the handicapped child's Goodenough MA was higher or lower than that of the control.

TABLE 2

Distribution of Majority Judgment by 9 Judges of 36 Matched Pairs on Right–Wrong and Goodenough Mental Age Variables

Goodenough MA	Majority Judgment		
	Right	Wrong	Total
Handicapped lower than control	21	2	23
Handicapped higher than control	2	11	13
Total	23	13	36

Note.—One of the matched pairs of children had equal Goodenough mental ages. Tables 2 and 3 do not include this pair.

From Table 2 it is apparent that the majority judgment was correct in 23 of the 36 possible cases, and that the Goodenough MA of the handicapped child was lower in 23 out of 36 cases, not significantly different from chance. (The *average* Goodenough MA difference was not significant at the .05 level.) It is evident also that there was a marked tendency for the majority judgment to be right when the Goodenough MA of the handicapped child was lower than that of the control. The majority judgment was usually wrong, however, when the Goodenough MA of the handicapped member of the pair was higher than that of his control. The chi square for the data in Table 2 was significant beyond the .001 level, after correction for continuity according to Yates. These tendencies are accentuated if only those cases were chosen where there was over two-thirds agreement among the judges. Table 3 shows the distribution of 26 "substantial agreement" cases.

Table 3 shows strikingly what has been the pattern of choices by the psychologist judges. In every case in the table, the judges (as a group) chose as the handicapped child the one with the lower Goodenough MA. In 18 out of 26 cases they were right. (It will be remembered that the average Goodenough MA of the controls was greater than that of the handicapped.) They were wrong in the eight

TABLE 3

DISTRIBUTION OF "SUBSTANTIAL AGREEMENT" JUDGMENTS ON RIGHT–WRONG AND GOODENOUGH MENTAL AGE VARIABLES

Goodenough MA	Group Judgment	
	Right	Wrong
Handicapped lower than control	18	0
Handicapped higher than control	0	8

cases where the handicapped child's Goodenough MA was *higher* than the control's.

Several of the judges gave as reasons for their decisions some evidence of self-depreciation, of a depressed self-concept in the protocols of some youngsters whom they therefore judged to be handicapped. A tendency to minify the Person was considered by some judges to be a projection of inferiority feelings arising as a result of the handicap. The height of the Person was measured as a check on this assumption. In 22 out of 37 cases, the Person drawn by the handicapped child was *larger* than that of the matched control child. This proportion is not significantly different from chance for $N = 37$.

DISCUSSION

The most likely inference is that the judges, while verbalizing psychodynamic bases for their decisions, were using primarily the intellectual characteristics of the drawings in judging which drawing belonged to the handicapped child and which to the control. They attributed the "better" drawing to the control and were right oftener than wrong, possibly because the Goodenough scores of the controls were higher in 62% of the cases than those of the handicapped children. If one took only the Goodenough MA scores and "judged" the better drawing to be that of the control, his percentage of correct judgments would be almost exactly the same as that of the total (or average) for the nine judges. It is possible, of course, that there are dynamic variables associated with "goodness" of the drawings, and the judges used these as criteria for their decisions. It is also possible that psychologists with intensive experience in the interpretation of drawings would have had a larger percentage of correct judgments, although the study of Schmidt and McGowan (1959) suggests otherwise.

SUMMARY

H-T-P drawings of 37 pairs of elementary school age children, one physically handicapped and one a control matched for age, sex, and IQ, were judged individually by nine psychologists to postdict which drawing was made by the handicapped child. The percentage of correct judgments was not significantly above chance for any one judge, although when the judgments were pooled the combined percentage correct was significantly greater than chance. There was a strong and significant tendency for the judges to attribute the drawing with the higher Goodenough score to the control subject. The handicapped child's drawing of a Person tended to be larger than that of the control, but the difference was not significant.

REFERENCES

BERREMAN, J. V. Implications of research in the social psychology of physical disability. *Except. Child.*, 1954, **20**, 347–350, 356–357.

BUCK, J. N. The H-T-P technique: A qualitative and quantitative scoring manual. *J. clin. Psychol.*, 1948, Monogr. Suppl. No. 5.

CRUICKSHANK, W. M. Psychological considerations with crippled children. In W. M. Cruickshank (Ed.), *Psychology of exceptional children and youth*. Englewood Cliffs, N. J.: Prentice-Hall, 1955.

FORCE, D. G., JR. Social status of physically handicapped children. *Except. Child.*, 1956, **23**, 104–107, 132–133.

GOODENOUGH, FLORENCE. *Measurement of intelligence by drawings*. New York: World Book, 1926.

LERNER, R., & MARTIN, M. What happens to the college student with a physical handicap? *Personnel guid. J.*, 1955, **34**, 80–85.

LEVY, J., & MICHELSON, BARBARA. Emotional problems of physically handicapped. *Except. Child.*, 1952, **18**, 200–206.

SCHMIDT, L. D., & MCGOWAN, J. F. The differentiation of human figure drawings. *J. consult. Psychol.*, 1959, **23**, 129–133.

SHERE, MARIE. Socio-emotional factors in families of the twin with cerebral palsy. *Except. Child.*, 1956, **22**, 197–199, 206–208.

WAWRZASZEK, F., JOHNSON, O. G., & SCIERA, J. L. A comparison of H-T-P responses of handicapped and nonhandicapped children. *J. clin. Psychol.*, 1958, **14**, 160–162.

WENAR, C. The effects of a motor handicap on personality: III. The effects on certain fantasies and adjustive techniques. *Child Develpm.*, 1956, **27**, 9–15.

WENAR, C. The degree of psychological disturbance in handicapped youth. *Except. Child.*, 1958, **25**, 7–10.

WHITEHOUSE, F. A. Habilitation: Concept and process. *J. Rehabilit.*, 1953, **19**, 3–7.

WRIGHTSTONE, J. W. Studies of orthopedically handicapped pupils. *Except. Child.*, 1957, **23**, 160–164, 176–177.

(Received May 23, 1960)

Journal of Consulting Psychology
1961, Vol. 25, No. 4, 288–292

CHANGES IN INTELLECTUAL FUNCTIONS OF CHILDREN IN A PSYCHIATRIC HOSPITAL

E. WESLEY HILER AND DAVID NESVIG
Mental Health Research Institute, Fort Steilacoom, Washington

It is a well-known fact that emotional disturbances can be intellectually incapacitating, interfering with concentration, learning, memory, judgment, and reasoning. Hence it is to be expected that such disturbances will not only interfere with academic functioning, but will also impair performance on psychological tests.

Despite the fact that IQ scores are widely interpreted as measures of intellectual capacity, the experienced clinician usually regards the test scores of seriously disturbed children as measures of intellectual functioning at the time of testing rather than as representing actual intellectual potential. The latter basic capacity or potential is usually inferred from those aspects of the test performance, such as the vocabulary level, which are assumed to be less influenced by emotional disturbance. Then, too, when a patient whose general test performance is poor or mediocre does well on some of the difficult items, one is led to suspect that the actual intellectual capacity is greater than the IQ score would suggest. Marked discrepancies among the Wechsler-Bellevue subtest scores are often the basis for inferring intellectual impairment of either functional or organic origins (Wechsler, 1958).

Although the average IQ of a group of normal children usually remains constant (Brown, 1950; Gehman & Matyas, 1956), certain individuals show a marked improvement and others a marked decline in test performance during the course of childhood. Such variations have been found to be related to emotional adjustment (Allen & Young, 1943; Clarke & Clarke, 1953; Despert & Pierce, 1946). It has also been reported that the IQ often rises as a consequence of successful psychotherapy (Chidester, 1934; Dulsky, 1942; Harrower, 1958; Hunsley, 1939; Miller, 1933) or other forms of treatment (Fisher, 1949; Markwell, Wheeler, & Kitzinger, 1953; Rabin, 1944). Change to a better environment also often leads to an improvement in test performance (Skeels & Harms, 1948). Children in warm, democratic homes were found to improve in IQ during childhood, whereas children in actively hostile and passive-neglectful homes tended to decline in IQ (Baldwin, Kalhorn, & Breese, 1945). The improvement in test performance is usually not uniform throughout the test. Certain aspects of test performance improve more than other aspects. Thus Harrower (1958) reports that a rise in Comprehension and Similarities scores on the Wechsler-Bellevue is related to clinical improvement in a group of adults. Kessler (1947) found that Picture Completion and Comprehension showed a significant rise after electroshock treatment, and Fisher (1949) reports significant improvement in Comprehension, Similarities, and Digit Symbol after EST.

The present study was carried out to investigate the effect on intellectual functions of the treatment program for children at Western State Hospital. This treatment program includes care by attendants selected for their ability to relate to children with warmth and with consistent discipline, classes conducted by teachers specially trained to deal with the emotionally disturbed, and participation in various recreational activities, arts, and crafts. An attempt is made to create a stable, noncompetitive environment with a minimum of stress. In addition, some patients receive tranquilizers and a few receive psychotherapy. Removal from an emotionally disturbing home environment and placement

in a relatively stable environment may be considered therapeutic in itself.

These children's principal intellectual deficiency was in Verbal IQ, which averaged about 14 points below Performance IQ. Therefore, we hypothesized that they would improve primarily on Verbal subtests and Verbal IQ. Inhibited, compulsively achieving children often have a Verbal IQ above the Performance IQ. In such children, one might expect Performance IQ to rise with clinical improvement. The problems of hospitalized children, however, differ from those of the typical neurotic patients seen in child guidance clinics. The child in a mental hospital is more apt to have a history of delinquent acting-out and school failure. His poorly controlled hostile impulses impede his learning in school, especially his learning of verbal material. He fails to acquire the normal fund of factual information, arithmetic skill, common sense, and reading ability which directly or indirectly is measured by the Verbal section of the Wechsler. Children with these characteristics are often sent to correctional schools. Several studies have reported the Verbal IQs of delinquents as below their Performance IQs (Bernstein & Corsini, 1953; Wechsler, 1958). A similar pattern was found for unsuccessful readers (Graham, 1952). The more acutely disturbed delinquent child is frequently sent to a state hospital. Many of the children in this hospital are either transferred to it from correctional schools or sent to the hospital as an alternative to correctional school.

Method

As part of another study, all children and adolescents up to the age of 18 who were admitted to Western State Hospital after a certain date were administered a battery of psychological tests at regular intervals. Out of the group of 40 cases which had been retested, we selected all those: (*a*) who had been given the Wechsler-Bellevue Form II upon admission, between 2 and 3 months after admission, and between 12 and 24 months after that; (*b*) who had not been out of the hospital more than 6 months between the second and third testing; (*c*) who had attended the hospital school. This left us with a sample of 20 cases. The average age of the sample on admission was 13.8 years with a range of 10.0 to 17.8 years. Diagnostic categories included three schizophrenics, three organics with behavior disorder, nine psychoneurotics, and five character disorders. Two additional cases, which had not had the first retest, were added to our sample for our comparisons of test improvement with rated improvement. Most of the cases in our sample had also been given the Bender-Gestalt test, from which was obtained a Pascal-Suttell *Z* score, the Goodenough Draw-A-Man, and the Gray Oral Reading Paragraphs Test.

On the Wechsler-Bellevue, in addition to making comparisons of IQs and subtest scores, subtests were grouped on the basis of Cohen's (1957, 1959) factor analysis of the Wechsler intelligence tests. Scores were obtained on four factors he found for the age level of our sample. The Verbal Comprehension Factor is the average of Information, Comprehension, Similarities, and Vocabulary; Perceptual Organization is the average of Object Assembly and Block Design; Freedom from Distractibility is the average of Arithmetic and Digit Span; and Judgment (Cohen's Verbal Comprehension II) is the average of Comprehension and Picture Completion.

In this study, it is assumed that changes in test scores after 1 or 2 months reflect practice effects or adjustment to the testing situation and to the hospital environment in general. The changes in scores after 12–24 months are assumed to reflect more basic changes in intellectual functioning.

Measures of behavioral changes were obtained in order to determine whether improvement in test performance was accompanied by actual improvement in condition. Change in condition was measured by a rating scale consisting of 20 variables, each on a five-point scale.

An overall rating of improvement was obtained on each patient by averaging the ratings of the specific traits. This procedure would only be justified if there were considerable homogeneity among the traits rated. A homogeneity coefficient was, therefore, computed by dividing the average between-trait (within subject) variance by the total variance, subtracting this from 1, and taking the square root. The coefficient was found to be .68; it indicates a moderate amount of homogeneity—sufficient to justify averaging the ratings on the individual traits to form a composite overall improvement score.

Each child was rated by staff members who were well acquainted with him. Altogether 25 raters were used, including 14 ward attendants, 5 school teachers, 4 social workers, 1 physician, and 1 EEG technician. The average number of raters for each child was 7.

An interrater reliability coefficient was computed by dividing the average between-rater (within subject) variance by the total variance, subtracting this from 1, and taking the square root. The coefficient was found to be .52. This is not very high. However, the use of many raters tends to counteract the unreliability of the individual raters and thus the mean ratings on each subject are believed to be sufficiently reliable to serve as measures of clinical improvement.

Different staff members rated different children; some tended to rate generally high, while others gave generally low ratings. Therefore, we measured the bias of each rater by comparing his overall-improve-

TABLE 1

CHANGES IN TEST SCORES ON RETEST

Variable	M_1	M_2	M_3	M_2–M_1	t	M_3–M_2	t
Full Scale IQ	86.70	90.15	89.70	3.45	2.46*	−.45	.27
Verbal IQ	81.15	81.50	82.45	.35	.24	.95	.65
Performance IQ	94.90	100.20	98.80	5.30	2.77*	−1.40	.61
Performance-Verbal IQ	13.75	18.70	16.35	4.95	2.27*	−2.35	.98
Subtest Total	79.45	83.30	88.00	3.85	2.20*	4.70	2.23*
Subtest *AD*	1.99	2.25	2.35	.26	2.87**	.10	.58
Information	4.70	4.80	5.55	.10	.11	.75	3.45**
Comprehension	6.60	6.10	7.40	−.50	1.00	1.30	3.83**
Digit Span	6.30	6.25	6.90	−.05	.08	.65	1.55
Arithmetic	4.40	4.55	5.10	.15	.44	.55	1.59
Similarities	6.85	7.55	7.35	.70	2.50*	−.20	.43
Vocabulary	6.95	6.85	6.75	−.10	.32	−.10	.32
Picture Arrangement	9.75	9.45	9.85	−.30	.52	.40	.77
Picture Completion	8.30	8.45	9.50	.15	.42	1.05	3.62**
Block Design	8.05	9.10	9.50	1.05	2.50*	.40	.84
Object Assembly	10.75	12.40	12.50	1.65	2.56*	.10	.17
Digit Symbol	6.80	7.80	7.60	1.00	2.55*	−.20	.30
Verbal Comprehension	6.25	6.33	6.75	.08	.41	.42	2.17*
Perceptual Organization	9.40	10.75	11.00	1.35	3.35**	.25	.56
Freedom from Distractibility	5.35	5.40	6.00	.05	.15	.60	1.92
Judgment	7.45	7.27	8.45	−.18	.47	1.18	5.23**
Bender Gestalt Z	99.55	95.60	84.40	−3.95	.73	−11.20	2.32*

Note.—M_1 is mean score on initial testing. M_2 is mean score on retest 2–3 months after first testing. M_3 is mean score on retest 12–24 months after second testing.

* $p < .05$.

** $p < .01$.

ment rating for each child with the overall-improvement rating for that child and taking the average of his deviations from the mean. We obtained the avererage rater bias for each child by averaging the bias of each of his raters. We then obtained corrected ratings for each child by subtracting the average rater bias from each child's average rating.

Our sample was small and not normally distributed on the psychological test variables; therefore, we dichotomized the scores on each variable at the median and used nonparametric statistics to compare test improvement with rated improvement. The significance level was evaluated with Fisher's exact test. The degree of relationship is indicated by the phi coefficient.

RESULTS AND DISCUSSION

Test Improvement of Group as a Whole

Table 1 contains the mean scores on the test variables on initial testing, retesting 2–3 months later, and retesting 12–24 months after that. It will be noted that on the first retest there is a rise of more than 5 points ($p = .05$) on the Performance IQ. This rise in Performance IQ results in a rise ($p = .05$) in the Full Scale IQ as well. There is no evidence of a rise in Verbal IQ. These results are consistent with other studies of practice effects (Derner, Aborn, & Canter, 1950; Hamister, 1949; Hays & Schneider, 1951; Steisel, 1951). Because of the increase in Performance IQ and the lack of increase in Verbal IQ the already large difference between Verbal and Performance IQ in this group becomes even larger ($p = .05$).

The following subtests show a significant improvement on the first retest: Object Assembly ($p = .05$), Block Design ($p = .05$), Digit Symbol ($p = .05$), and Similarities ($p = .05$). The only factor to show a rise on the first retest is Perceptual Organization ($p = .01$). These changes are undoubtedly due at least in part to practice effect. Object Assembly, Block Design, Digit Symbol, and Picture Arrangement were reported as showing the greatest amount of practice effect after 1 and 4 weeks for a group of normals (Derner et al., 1950). Some of the improvement may reflect the stabilizing effect that the hospital has on these patients during the first few months. This stabilization would reduce loose associations and result in an improvement in the ability to perceive relationships as measured by the Similarities subtest.

Because of the marked increase on certain subtests which are already high for this group, there is an increase in subtest variability ($p = .01$). This suggests that one should be cautious in making inferences of pathology on the basis of the subtest variability of patients who have been tested before. Because practice has a greater effect on the Performance IQ than on the Verbal IQ, one would usually expect to find Performance IQ higher than Verbal IQ for individuals who have been tested previously.

After a period of 12–24 months there is an appreciable rise in the subtest total ($p = .05$), but this does not result in an improvement in IQ because by that time the children fall in a different age bracket and require a higher subtest total to achieve the same IQ. After 12–24 months the greatest rise occurs on the Information, Comprehension, and Picture Completion subtests ($p = .01$).

The scores on two of Cohen's factors show a significant rise after 12–24 months. The improvement in the Judgment factor is highly significant ($p = .01$). The improvement on the Verbal Comprehension factor was smaller ($p = .05$). A small rise on the Freedom from Distractibility factor approaches significance.

It is interesting to note that the subtests and factors showing the most improvement after 12–24 months were not the ones showing improvement after 2–3 months.

Test Improvement Related to Ratings of Improvement

There is a significant relationship ($\Phi = .55$, $p = .05$) between improvement in Verbal IQ and ratings of general improvement. Children showing more than the median amount of improvement went up 4.8 points in Verbal IQ, whereas children who showed less than the median amount went down 1.7 points.

The relationship between rated improvement and Full Scale IQ was smaller but was also significant ($\Phi = .36$, $p = .05$).

There was no significant relationship between clinical improvement and improvement in Performance IQ ($\Phi = .10$, $p = ns$).

Improvement on only one subtest, Digit Span, was significantly related to overall clinical improvement ($\Phi = .46$, $p = .05$). Information and Comprehension approached significance ($p = ns$).

Changes on the Goodenough IQ and the Bender-Gestalt did not seem to be related to clinical improvement. Improvement on the Gray reading test was, however, related to rated improvement ($N = 12$, $\Phi = .66$, $p = .05$). Those rated as improving more than the median amount went up 1.5 years in reading level while those improving less than the median amount went up only .6 years.

A comparison of each of the 20 rated variables and the psychological test variables showed the following significant relationships ($p = .05$ unless specified):

1. Improvement in Verbal IQ was related to improvement in most of the 20 clinical variables but only the following relationships were statistically significant beyond the .05 level of confidence: Achievement in School, Development of New Interests and Goals, Ability to Concentrate and Resist Distractions, Reduction in Anxiety.

2. Achievement in School and Development of New Interests and Goals were significantly related to improvement on the Information and Comprehension subtests. Reduction in Anxiety was significantly related to improvement in Digit Span. Improvement in Dependability was related to improvement on the Arithmetic subtest ($\Phi = .55$, $p = .05$). Improvement in Ability to Concentrate and Resist Distractions was related to improvement in Digit Span and the Gray reading test ($\Phi = .56$, $p = .05$).

3. Improvement on the Digit Symbol subtest was significantly related to Decrease in Bizarre Thought Processes and also to Decrease in Delinquent Tendencies.

Summary

This study was carried out to determine the effect of a hospital program on the intellectual functioning of emotionally disturbed children.

A sample of 20 children with a mean age of 13.8 years was tested on admission to the hospital, retested 2–3 months later, and retested again 12–24 months after that.

A significant rise occurred after 2–3 months on Wechsler Performance IQ, Full Scale IQ, subtest total, subtest average deviation, Similarities, Block Design, Object Assembly, and Digit Symbol, and the Perceptual Organization factor. This improvement is partially at-

tributable to practice effect, but the improvement in Similarities perhaps reflects a decrease in bizarre or irrelevant thought processes.

After 12–24 months, a marked improvement was shown on the Information, Comprehension, and Picture Completion subtests; the Judgment and Verbal Comprehension Factors; and Bender-Gestalt performance. The group as a whole appears to be better organized perceptually, to have more common sense, better judgment, and an increased ability to perceive relationships and distinguish between essential and unessential aspects of a situation.

The Performance, Verbal, and Full Scale IQs did not go up for the group as a whole. However, it was noted that certain patients did improve considerably on these scales while others declined. It was hypothesized that such differences in the direction of change would be related to improvement or deterioration of the patient's condition.

Ratings of improvement on 20 clinical variables were obtained from the hospital staff. Improvement in Verbal IQ, Full Scale IQ, Digit Span, and Gray Oral Reading test level was significantly related to overall clinical improvement. Improvement in IQ and in specific subtests was found to be related to improvements on specific traits.

It is concluded that most hospitalized children have problems which cause them to be retarded in verbal skills; as they improve, their Verbal IQ rises. It is suggested that the initial Verbal IQ is not a fair estimate of the intellectual capacities of emotionally disturbed children, and that Performance IQ may provide a more accurate measure of these children's actual intelligence.

REFERENCES

ALLEN, M. E., & YOUNG, F. M. The constancy of the intelligence quotient as indicated by retests of 130 children. *J. appl. Psychol.*, 1943, **27**, 41–60.

BALDWIN, A. L., KALHORN, J., & BREESE, F. H. Patterns of parent behavior. *Psychol. Monogr.*, 1945, **58**(3, Whole No. 268).

BERNSTEIN, R., & CORSINI, R. Wechsler-Bellevue patterns of female delinquents. *J. clin. Psychol.*, 1953, **9**, 176–179.

BROWN, G. L. On the constancy of the IQ. *J. educ. Res.*, 1950, **44**, 151–153.

CHIDESTER, L. Therapeutic results with mentally retarded children. *Amer. J. Orthopsychiat.*, 1934, **4**, 464–472.

CLARKE, A. D. B., & CLARKE, A. M. How constant is the IQ? *Lancet*, 1953, **265**, 877–880.

COHEN, J. The factorial structure of the WAIS between early adulthood and old age. *J. consult. Psychol.*, 1957, **21**, 283–290.

COHEN, J. The factorial structure of the WISC at ages 7–6, 10–6, and 13–6. *J. consult. Psychol.*, 1959, **23**, 285–299.

DERNER, G. F., ABORN, M., & CANTER, A. H. The reliability of the Wechsler-Bellevue subtests and scales. *J. consult. Psychol.*, 1950, **14**, 172–179.

DESPERT, J. L., & PIERCE, H. O. The relation of emotional adjustment to intellectual function. *Genet. psychol. Monogr.*, 1946, **34**, 3–56.

DULSKY, S. G. Affect and intellect: An experimental study. *J. gen. Psychol.*, 1942, **27**, 199–220.

FISHER, K. A. Changes in test performance of ambulatory depressed patients undergoing electroshock therapy. *J. gen. Psychol.*, 1949, **41**, 195–232.

GEHMAN, ILA H., & MATYAS, B. P. Stability of the WISC and Binet tests. *J. consult. Psychol.*, 1956, **20**, 150–152.

GRAHAM, E. E. Wechsler-Bellevue and WISC scattergrams of unsuccessful readers. *J. consult. Psychol.*, 1952, **16**, 268–271.

HAMISTER, R. The test-retest reliability of the Wechsler-Bellevue intelligence test (Form I) for a neuropsychiatric population. *J. consult. Psychol.*, 1949, **13**, 39–43.

HARROWER, MOLLY. *Personality change and development as measured by projective techniques.* New York: Grune & Stratton, 1958.

HAYS, W., & SCHNEIDER, B. A test-retest evaluation of the Wechsler Forms I and II with mental defectives. *J. clin. Psychol.*, 1951, **7**, 140–143.

HUNSLEY, Y. L. Intelligence, as reflected by work habits, attitude and behavior, does change. *Sch. Soc.*, 1939, **50**, 682–684.

KESSLER, LUCILLE. Intellectual changes in schizophrenic patients following electroshock therapy. Unpublished master's thesis, New York University, 1947.

MARKWELL, E. D., JR., WHEELER, W. M., & KITZINGER, HELEN. Changes in Wechsler-Bellevue test performance following prefontal lobotomy. *J. consult. Psychol.*, 1953, **17**, 229–231.

MILLER, E. Emotional factors in intellectual retardation. *J. ment. Sci.*, 1933, **79**, 614–625.

RABIN, A. I. Fluctuations in the mental level of schizophrenic patients. *Psychiat. Quart.*, 1944, **18**, 78–92.

SKEELS, H. M., & HARMS, IRÈNE. Children with inferior social histories: Their mental development in adoptive homes. *J. genet. Psychol.*, 1948, **72**, 283–294.

STEISEL, I. M. The relation between test and retest scores on the Wechsler-Bellevue scale (Form I) for selected college students. *J. genet. Psychol.*, 1951, **79**, 155–162.

WECHSLER, D. *The measurement and appraisal of adult intelligence.* Baltimore: Williams & Wilkins, 1958.

(Received June 6, 1960)

Journal of Consulting Psychology
1961, Vol. 25, No. 4, 293–295

A COMPARISON OF SOCIAL AND SOLITARY MALE DELINQUENTS[1]

MARY H. RANDOLPH, HAROLD RICHARDSON, AND RONALD C. JOHNSON
San Jose State College

Most contemporary theories concerning the causes of juvenile delinquency might be roughly differentiated into two areas of emphasis: psychologically oriented theories, such as those expressed by the Gluecks (1950), Healy and Bronner (1936), and Lindner (1944); and sociologically oriented theories, such as those expressed by Shaw and McKay (1942), Sutherland (1955), Thrasher (1936), and others. Both points of view may be defended since it seems likely that psychological and sociological forces interact to varying degrees in the histories of most delinquents.

Lindesmith and Dunham (1941) have differentiated between the socialized and the individualized criminal. They state that the socialized criminal is one who commits crimes that are supported and prescribed by his culture, so that, by committing a crime, the criminal gains in status and recognition. The socialized delinquent or criminal acts in close collaboration with other persons and is dependent on them for the continuation of his criminal career. The individualized criminal, on the other hand, acts for reasons that are personal and private. He commits his crimes alone and, in theory, is a stranger to others who commit similar crimes. His criminal act is not an acceptable form of behavior in his social milieu. The socialized criminal seems likely to be a rather normal person, in a psychological sense, who holds deviant social values common to his subcultural group. The individualized criminal, at odds with his own primary groups, seems likely to be an individual whose criminality is merely symptomatic of deeper psychological pressures. Johnson (1949) has suggested that the solitary delinquent is an individual with a "conscience defect" unconsciously fostered by the parents, while the social delinquent is the product of a subculture with delinquent values. Bloch and Flynn (1956) made similar statements. Within this framework, the sociological theories would seem most useful in explaining the delinquency of male social delinquents—gang members and others who commit delinquent acts in the company of others, while psychological explanations might best account for the individualized or solitary delinquent.

It has been found (Hewitt & Jenkins as reported by Wattenberg & Balistrieri, 1950) that juvenile gang members are likely to come from homes of a lower socioeconomic stratum, while nongang members showed more indications of coming from stressful or depriving homes of a middle socioeconomic level. Johnson (1950) found that solitary delinquents were far more often recidivists than were social delinquents, even though the majority of his social delinquents were members of well organized juvenile gangs. This finding might be expected if individual or solitary delinquents are merely acting out symptoms of deep-seated and unresolved psychological stresses.

Beyond these scanty data, little is known about genotypic or phenotypic variation between solitary and social delinquents, even though treatment techniques used for the two groups might be made more effective if this information were available. The purpose of this study is to compare solitary and social delinquents with regard to several "sociological" and "psychological" variables.

METHOD

Subjects

The sample consisted of 62 boys, aged 14–18, who had been adjudged legally as juvenile delinquents. Of these, 52 subjects were at a "ranch" for delinquent boys and the other 10 were in custody, await-

[1] This report is based on a thesis submitted (by the first author) to the Department of Psychology, San Jose State College, January 1960.

TABLE 1

NUMBER OF SOLITARY AND SOCIAL DELINQUENTS FROM EACH SOCIOECONOMIC LEVEL

Delinquent	Upper Middle	Lower Middle	Upper Lower	Lower Lower	*N*
Solitary	4	7	5	2	18
Social	1	4	15	19	39

Note.—$\chi^2 = 15.83$, $p < .01$.

ing placement at this ranch. All boys were of at least dull normal intelligence. Of the original sample, one subject was eliminated because of insufficient responses to test items, and four subjects were eliminated because extensive further examination showed them to have had mixed (solitary and social) delinquent careers. Fifty-seven subjects remained. Of these subjects, 39 had always been social and 18 had always been solitary in their known delinquencies.

Measuring Devices

Each subject was administered a Wechsler Adult Intelligence Scale (WAIS), and a Minnesota Multiphasic Personality Inventory (MMPI). Socioeconomic status was determined by a local adaptation (Hodges, unpublished) of the Warner Index (Warner, 1949).

Procedure

All subjects were tested inside an institutional setting. Tests were administered and scored according to standardized procedure except that all questions on the MMPI were read aloud to subjects, while the subjects read the questions in the booklet, in order to minimize difficulties in comprehension. All subjects knew that test results were confidential and would not influence placement. Only two MMPI records (both of social delinquents) had to be discarded as invalid.

RESULTS

On the WAIS IQ scores the social delinquents had a mean of 93.23 with a standard deviation of 9.14. The solitary delinquents had a mean of 105.00 with a standard deviation of 11.19. The *t* test of differences between these means was significant beyond the .01 level. The *F* test of differences between variances was not significant. The mean IQ score of the solitary delinquents was exceeded by only 15% of the social delinquents.

Many more solitary delinquents came from upper socioeconomic levels than did social delinquents, as shown in Table 1.

MMPI profiles are presented in Figure 1.

Mean differences between social and solitary delinquents on the validating scales *L*, *F*, and *K* were not significant. Both groups scored rather high on *F*, found to be an indicator of psychopathology (Kazan & Scheinberg, 1945; Modlin, 1956) and of delinquency (Hathaway & Monachesi, 1953). Profiles of the two groups are similar but solitary delinquents, as a group, appear somewhat more disturbed. All mean differences for the diagnostic scales were significant except for the *Ma* scale. Differences on *Mf*, *Pa*, and *Si* were significant beyond the .05 level. Differences on *Hs*, *D*, *Hy*, *Pd*, *Pt*, and *Sc* were significant beyond the .01 level of confidence. The code type of the solitary delinquent was 8479'612305–, (4, 13, 11). The social delinquent's code type was '8497613–, (3, 11,

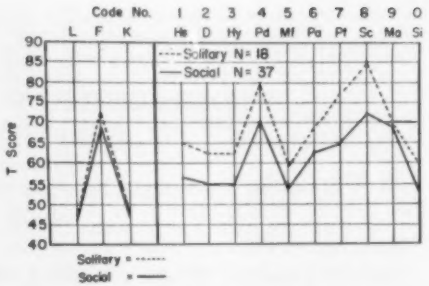

FIG. 1. Mean profiles for social and solitary delinquents on the MMPI.

TABLE 2

DATA COMPARING SOCIAL AND SOLITARY DELINQUENTS ON THE MMPI

(Social $N = 37$, Solitary $N = 18$)

Scale	Mean Social	Mean Solitary	*SD* Social	*SD* Solitary	*t*
L	3.45	3.55	2.42	2.04	1.51
F	11.23	12.76	4.58	4.55	1.49
K	10.15	10.94	3.97	6.62	.68
Hs	14.03	17.05	3.10	4.03	2.97**
D	18.77	21.94	3.98	4.71	3.23**
Hy	19.21	23.33	4.35	4.47	3.10**
Pd	27.62	30.89	4.38	4.63	3.05**
Mf	22.18	25.38	4.18	4.71	2.34*
Pa	12.36	14.38	2.95	3.27	2.22*
Pt	30.26	35.55	5.95	6.29	3.37**
Sc	33.41	39.83	7.61	7.93	3.54**
Ma	24.18	25.05	4.17	4.70	.78
Si	28.85	32.55	6.09	8.27	2.48*

Note.—These figures are expressed in MMPI raw scores.
* Significant at .05 level.
** Significant at .01 level.

10). Although both types had high excitors (*Pd, Sc, Ma*), only the solitaries had high suppressors (*Si, D, Mf*), presumably indicating neurotic trends. Complete statistical data for comparing social and solitary delinquents on the MMPI are given in Table 2.

Discussion

It seems relatively clear, from these results, that solitary and social delinquents differ considerably from each other. The solitary delinquent seems likely to come from a higher socioeconomic level and to be of higher intellectual ability than the social delinquent, but to be considerably more maladjusted. These findings might explain why, once embarked upon a course of delinquent behavior, the solitary delinquent is more inclined to be a recidivist.

Differences between the two groups of delinquents might be taken into account in diagnosis, prognosis, and treatment. The prognosis for the solitary delinquent without some form of therapy seems likely to be poorer than for social delinquents. Current sociologically oriented treatment techniques might more often be sufficient for the rehabilitation of the social delinquent. So long as therapy services are in short supply, a more economical use of therapists' time might result from a consideration of the obtained differences between social and solitary delinquents.

Summary

Test data were obtained from 57 delinquent boys, aged 14–18, all of whom, at the time of testing, were within an institutional setting. Thirty-nine of the subjects were "social delinquents" who committed their crimes in the company of others. Eighteen subjects were "solitary delinquents" who had committed their delinquencies alone. Wechsler intelligence test (WAIS) scores indicated that solitary delinquents are significantly higher than social delinquents in intellectual ability. Solitary delinquents were also significantly higher in socioeconomic status than were social delinquents. MMPI profiles of the two groups were similar, but with a significantly greater elevation in all scales except *Ma* for the solitary delinquent. The solitary delinquent appears more likely to be a psychologically deviant individual who comes from an ostensibly normal environment, while the social delinquent seems far less deviant, in a psychological sense, but comes from an environment where certain sociological factors, presumably causal to delinquency, are operating. Certain implications of these findings were discussed.

References

Bloch, H. A., & Flynn, F. T. *Delinquency*. New York: Random House, 1956.

Glueck, S., & Glueck, Eleanor T. *Unraveling juvenile delinquency*. New York: Commonwealth Fund, 1950.

Hathaway, S. R., & Monachesi, E. D. *Analyzing and predicting juvenile delinquency with the MMPI*. Minneapolis: Univer. Minnesota Press, 1953.

Healy, W., & Bronner, Augusta F. *New light on delinquency and its treatment*. New Haven: Yale Univer. Press, 1936.

Johnson, Adelaide M. Juvenile delinquency. In S. Arieti (Ed.), *American handbook of psychiatry*. Vol. 1. New York: Basic Books, 1949.

Johnson, R. C. Causal factors in the delinquency of fifty Denver boys. Unpublished master's thesis, University of Denver, 1950.

Kazan, A. T., & Sheinberg, I. M. Note on the significance of the validity score (*F*) in the MMPI. *Amer. J. Psychiat.*, 1945, **102**, 181–183.

Lindesmith, A. R., & Dunham, H. W. Some principles of criminal typology. *Soc. Forces*, 1941, **19**, 307–314.

Lindner, R. *Rebel without a cause*. New York: Grune & Stratton, 1944.

Modlin, H. C. A study of the MMPI in clinical practice. In G. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.

Shaw, C. R., & McKay, H. D. *Juvenile delinquency in urban areas*. Chicago: Univer. Chicago Press, 1942.

Sutherland, E. H., & Cressey, D. R. *Principles of criminology*. Philadelphia: Lippincott, 1955.

Thrasher, F. M. *The gang*. Chicago: Univer. Chicago Press, 1936.

Warner, W. L., Meeker, M., & Eels, K. *Social class in America*. Chicago: Science Research Associates, 1949.

Wattenberg, W. W., & Balistieri, J. J. Gang membership and juvenile misconduct. *Amer. sociol. Rev.*, 1950, **15**, 744–752.

(Received June 6, 1960)

Journal of Consulting Psychology
1961, Vol. 25, No. 4, 296–303

THE DIMENSIONALITY OF RATINGS OF THERAPIST VERBAL RESPONSES[1]

EDMUND S. HOWE AND BENJAMIN POPE
University of Maryland School of Medicine

The last few years have witnessed an important trend in research in psychotherapy, toward study of the therapist as an independent variable in the dyadic relationship. This trend has in part shifted the focus of empirical attention away from such issues as theoretical differences per se among "schools" of psychotherapy, and has instead directed research toward rigorous quantification of basic variables cutting across theoretical and practical divergences among therapists. Among the most penetrating studies of this kind are those investigating the dimension and the dimensionality of Depth of Interpretation (e.g., Harway, Dittmann, Raush, Bordin, & Rigler, 1955; Raush, Sperber, Rigler, Williams, Harway, Bordin, Dittmann, & Hayes, 1956; Speisman, 1959). Other investigators have approached presumedly different aspects of therapist verbal behavior such as Directiveness (e.g., Snyder, 1953) and Ambiguity (e.g., Osburn, 1951), to mention but two.

Howe and Pope (1961) and Pope, Howe, and Finesinger (1959) have more recently approached the problem of therapist verbal behavior from the standpoint of Finesinger's (1948) enunciation of a principle of Minimal Activity. Using attributes of "Ambiguity," "Lead," and degree of "Inference" as aspects of the concept of Therapist "Activity," an Activity Scale was constructed on the basis of ratings, by psychiatrists, of a representative sample of 50 therapist verbal responses. The order of reliability observed among ratings used in constructing the Activity Scale and ratings obtained from *application* of the scale was about .50. While this average reliability coefficient compares quite favorably with those reported by earlier investigators of a different aspect of therapist verbal behavior (e.g., Harway et al., 1955; Raush et al., 1956) 75% of the total variance is nevertheless left as unexplained "error." The rather low reliability indices obtained in these and other studies is quite possibly due to the assumed "dimension" in rating studies being multiple, rather than unitary; as Raush et al. (1956) earlier pointed out. For, as Coombs (1951) and Bordin, Cutler, Dittmann, Harway, Raush, and Rigler (1954) have written, it is quite possible to "force" unidimensionality even where it does not mathematically exist. Indeed, even while the studies of therapist activity were being performed, it became subtly obvious to the experimenter that the resistance of some of the psychiatrists to performance of the rating task devoid of a flesh-and-blood patient was at least in part due to an apparent confounding of their *evaluative* attitudes with their judgments of activity according to relative ambiguity, lead, and inference involved in each therapist response. Were this the case, then one would statistically predict relatively low reliability among ratings along a one-dimensional continuum.

The study reported here was thus undertaken to explore the dimensionality of ratings of such types of therapist verbal responses as

[1] This paper arose out of research supported by Pilot Evaluation Grant No. 2M-6408 from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service. The late Jacob E. Finesinger was the principal investigator. Thanks are acknowledged for his continuous encouragement and wholehearted support of this work until his untimely death in June 1959. Final completion of the work and of the present manuscript was indirectly facilitated by Research Grant M-3355 (also from the National Institute of Mental Health) of which ESH is the principal investigator. A paper based upon this research was presented to Division 12 at the Annual Convention of the American Psychological Association in Chicago, September 1960.

had been used in the earlier presumptive one-dimensional studies of therapist activity. Application of a 40-scale semantic differential to therapist responses thus facilitated a check on the two general propositions that such ratings of therapist verbal responses would be primarily of an evaluative nature, and at least two-dimensional.

Method

Raters. A decision was made to solicit the services of Board-certified or Board-eligible psychiatrists, rather than of psychiatrists having had some minimum amount of therapeutic experience. The subjects were drawn from the Psychiatric Institute at the University of Maryland School of Medicine, from those engaged in full-time private practice in Baltimore City, from the National Institute of Mental Health, from the Walter Reed Army Hospital and Institute for Research, from Chestnut Lodge, and from Spring Grove State Hospital, Maryland. The booklets described below were mailed, after verbal agreement by telephone, to 50 subjects. Of the 38 booklets returned,[2] 3 were discarded because of inadvertant omission of at least 1 page, either by experimenter or by subject. Data from the remaining 35 subjects were analyzed.

Therapist Verbal Responses. Ten therapist responses from the set of 50 used in the earlier rating study were selected for experimentation. These responses, each chosen for specific reasons of either close a priori similarity to or extreme a priori dissimilarity from at least one other response, are presented in Table 1. These reasons are now briefly summarized. Responses 1 and 10 were selected for study because they were originally given the most extensive mean Activity Level (AL) ratings. Responses 2 and 3 were selected because they were originally rated equally, and were clearly of a very low-active, facilitating nature. Responses 4 and 5 were likewise chosen since both were about equally and yet somewhat more specifically focused than 2 and 3, and both still quite low-active. Response 6 is unique in the set of 10 studied here, since it is a highly specific, *objective* question. Responses 7 and 8 were included since they constitute interpretive responses of subjectively equal depth, and have equal ALs. Response 9 is evidently a reassurance/supportive type of response, subjectively quite distinctive in its own right.

Since the 10 responses could clearly be used as "markers," the potential disadvantage of making a deliberate selection (with the attendant risk of bias) was largely offset by the definitive role the responses could play in the face validation of certain empirical outcomes to be described later. The question whether Response 10 (and perhaps even Response 9) normally occur too infrequently to be included without risk of seriously biasing the empirical outcomes in this type of research can not be fully dealt with here, but it deserves comment. Response 10 ("You have to tell me whether it is so or not") is essentially a demanding and persuasive operation toward the patient. As such, it is largely a tabooed response, negatively valued by most therapists. (A comparable, though weaker argument might be made by some therapists with regard to Response 9, which is essentially a reassurance operation instigated by the therapist himself.) Operations of a persuasive nature may, however, take many and variegated forms. Persuasion qua persuasion rarely occurs in most accepted psychotherapies. But the type of persistence and hounding of the patient's thoughts that occurs in the published interviews of Deutsch and Murphy (1955), to cite only one example, reflects from an operational standpoint a respectable unwillingness of the therapist to let the patient "get away" until he reports that which the therapist, perhaps unconsciously, wants to hear. Thus, if the concept of persuasiveness be regarded more as an "attitude of mind" in the therapist, than as the manifest form that his communications to the patient actually take, then it becomes more reasonable to include some sort of verbal stimulus *connoting* such a therapist attitude. Indeed, inclusion of such an anchor stimulus would, under the foregoing conditions, be as like to *reduce* as to foster bias in the empirical outcome.

TABLE 1

The 10 Therapist Verbal Responses Presented to the Raters and Original Mean Activity Levels

No.	Therapist Verbal Response	AL
1	Hm-hm	1.4
2	And —?	2.4
3	What's been happening?	2.6
4	By *brooding* you mean —?	4.3
5	Your *heart* —?	4.4
6	How much do you earn?	6.2
7	Perhaps he feels attracted to you	7.8
8	Maybe you participate in this, encouraging them to depend on you more than you think.	7.8
9	I hope you don't think I'm impatient with you, because I'm not	8.7
10	You have to tell me whether it is so or not.	9.7

Note.—Based on the ratings of 30 psychiatrists during the first, 50-item rating study.

[2] Many thanks are here expressed to all of those unnamed persons who were kind enough to perform the ratings.

TABLE 2

THE SEVEN SETS OF BIPOLAR, ADJECTIVAL SCALES

Hypothetical Variable	Bipolar Scale
Ambiguity	Spacious-Constricted
	Colorless-Colorful
	General-Specific
	Unfocused-Focused
	Commonplace-Unique
	Vague-Precise
Stressfulness	Cold-Warm
	Calm-Excitable
	Sober-Drunk
	Relaxed-Tense
	Cautious-Rash
	Relieving-Painful
Inference	Subtle-Obvious
	Inferential-Logical
	Intuitive-Rational
	Deep-Shallow
	Private-Public
Lead	Following-Leading
	Accepting-Demanding
	Conforming-Directing
Evaluative	Skillful-Unskillful
	Reputable-Disreputable
	Wise-Foolish
	Good-Bad
	Accepting-Rejecting
	Sensitive-Insensitive
	Acceptable-Unacceptable
	Valuable-Worthless
Activity	Still-Vibrant
	Static-Dynamic
	Slow-Fast
	Inert-Energetic
	Passive-Active
Potency	Muted-Blatant
	Weak-Strong
	Soft-Hard
	Thin-Thick
	Far-Near
	Small-Large
	Dull-Sharp

Note.—The classification is arbitrary in several cases.

The Rating Booklet. Each rater was presented with a 20-page booklet. Each successive pair of pages contained a total of 40 seven-point, bipolar, adjectival scales, the therapist response to be judged appearing at the top of each pair of pages. The 40 scales were of course the same for each response, and they appeared in the same order. The order in which the therapist responses appeared, however, was varied in four ways (viz.: Numbers 1–10; Numbers 10–1; Numbers 6–10, 1–5; Numbers 5–1, 10–6). In all other respects the format of the instructions to the subject followed that described by Osgood, Suci, and Tannenbaum (1957). The subject was instructed to assume that each response was made during an initial interview. It was considered desirable to permit the subject to project his own feelings about context, since the generality of the findings would thereby be enhanced.

The Adjectival Scales. The set of 40 scales was selected after an exhaustive examination both of *Roget's Thesaurus,* and of published work (e.g., Osgood et al., 1957) using different forms of the Semantic Differential. It was decided to include scales having some connotative reference to Ambiguity, Lead, and Inference, since these attributes had been used earlier as rough working referents of the concept of Activity. Scales having connotative reference to Stressfulness were also included because of such frequent explicit claims by previous subjects as, "I would consider something more 'Active' if it tends to upset the patient." Finally, in view of their ubiquitous appearance in numerous reported studies, scales were included having established relevance to Osgood's Evaluative, Potency, and Activity dimensions. These seven sets of bipolar adjectives are presented in Table 2. The present classification of individual items into the seven sets is, of course, purely arbitrary in several instances.

Treatment of Data. A 40×40 intercorrelation matrix, with $N = 35$ (Judges) $\times$ 10 (Verbal Responses) $= 350$ pairs entering into each correlation, was obtained by IBM. The matrix was factored by the complete centroid method of Thurstone (1947), and since a maximum of seven hypothetical dimensions was implied by the initial categorization of the scales into seven sets, a total of nine factors was extracted, of which three were significant and interpretable. The centroid was orthogonally rotated by the quartimax method (Neuhaus & Wrigley, 1954).

RESULTS

The Factor Analysis

The rotated factor loadings are presented in Table 3. The first factor to emerge accounts for 33% of the total variance and 60% of the common variance. It is most clearly defined by the bipolar scales foolish-wise, unacceptable-acceptable, skillful-unskillful, good-bad, and valuable-worthless. The first two scales have rather pure, high negative loadings and the last three rather pure, high positive loadings on the first factor. Other, somewhat less pure but still significant loadings are observed for scales tense-relaxed, blatant-muted, and rejecting-accepting (negatively

TABLE 3

ROTATED FACTOR LOADINGS

Scale	I	II	III	h^2
Tense-Relaxed	−.83	−.16	.03	.72
Skillful-Unskillful	.87	−.02	−.04	.76
Hard-Soft	−.78	−.26	−.06	.68
Passive-Active	.38	.65	−.03	.57
Near-Far	.52	−.34	.01	.39
Reputable-Disreputable	.80	−.01	−.03	.64
Spacious-Constricted	.60	.11	.13	.39
Foolish-Wise	−.90	.02	.10	.82
Cautious-Rash	.83	.27	−.03	.76
Blatant-Muted	−.83	−.28	−.08	.77
Colorless-Colorful	.09	.77	−.09	.61
Still-Vibrant	.17	.77	−.10	.63
Leading-Following	−.25	−.52	.04	.33
Small-Large	−.13	.54	−.09	.32
Accepting-Demanding	.75	.16	.20	.63
Strong-Weak	.15	−.73	−.05	.56
Specific-General	−.20	−.69	−.13	.53
Good-Bad	.90	−.04	−.05	.81
Rejecting-Accepting	−.82	.01	−.11	.68
Static-Dynamic	−.32	.60	−.17	.49
Conforming-Directing	.38	.46	.07	.36
Thick-Thin	−.11	−.48	.10	.25
Focused-Unfocused	−.17	−.63	−.06	.43
Slow-Fast	.19	.67	−.12	.50
Obvious-Subtle	−.61	−.14	−.30	.48
Sensitive-Insensitive	.84	−.10	.14	.74
Calm-Excitable	.83	.18	.00	.72
Energetic-Inert	−.16	−.82	.02	.70
Cold-Warm	−.61	.27	−.09	.45
Inferential-Logical	.16	−.17	.74	.60
Unique-Commonplace	−.08	−.38	.28	.23
Vague-Precise	.07	.71	.14	.53
Unacceptable-Acceptable	−.88	.00	.08	.78
Relieving-Painful	.71	.11	.09	.52
Sharp-Dull	−.27	−.41	.05	.24
Sober-Drunk	.51	−.21	−.14	.32
Shallow-Deep	−.61	.37	−.11	.52
Private-Public	.37	−.26	.16	.23
Intuitive-Rational	.24	−.05	.78	.67
Valuable-Worthless	.89	−.09	−.08	.81
	$\Sigma a^2 = 13.38$	7.15	1.66	$\Sigma h^2 = 22.19$

loaded), and for scales cautious-rash, sensitive-insensitive and calm-excitable (positively loaded). There is no question but that the first factor may be appropriately interpreted as one of Professional Evaluation. Its nature implies that the "good" therapist, the one who is thoroughly reputable and skilled, uses responses which are cautious, relaxed, muted, accepting, sensitive, and calm. The emergence of such a factor *first*, further implies, as we shall see, that such evaluative connotations of the verbal responses are more compelling, more salient than are the connotations of the second and subsequent factors interpreted.

The second factor accounts for 18% of the total variance and 32% of the common variance. It is most clearly defined by scales colorless-colorful, still-vibrant, and vague-precise (positive loadings), and by scales energetic-inert and strong-weak (negative loadings). Other, somewhat less pure but still significant loadings on the second factor are ob-

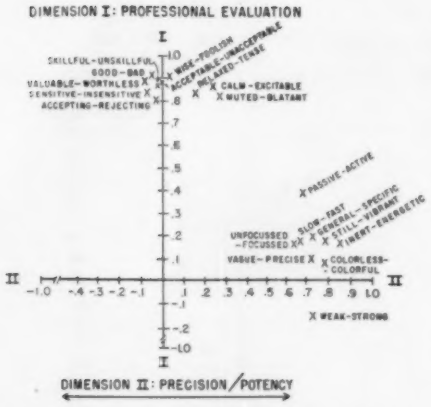

FIG. 1. Scales having significant rotated loadings on Factor 1, or on Factor 2.

served for scales passive-active and slow-fast (positive loadings) and for scales specific-general and focused-unfocused (negative loadings). The second factor is interpreted as one of Precision/Potency, although Ambiguity/Passivity would be almost as appropriate a label. This factor clearly refers to those attributes of therapist behavior variously referred to as Activity, Ambiguity, and the like. Its nature is reminiscent of the "dynamism factor" so labeled by Osgood et al. (1957) to describe the apparent coalescence of their second (Activity) and third (Potency) factors in the judgment of sociopolitical concepts.

The third factor, which accounts for only a little over 4% of the total variance and 7% of the common variance, is represented by only two scales: inferential-logical and intuitive-rational, both being quite highly positively loaded on this factor. The rotated loadings for these scales are, respectively, .74 and .78, and they are fairly pure scales; but there are no other scales even approaching significant loadings on this factor. Consequently, it is rather difficult to interpret this factor with great conviction; for it is probably more a factor of Subjectivity/Objectivity rather than one of inference. The difficulty in interpreting this factor reflects in part a selection of scales of which the referents, retrospectively considered, are somewhat nebulous. The fourth through sixth factors are all significant ($p < .05$) according to Humphreys' Rule (Fruchter, 1954, pp. 79–80), the respective percentages of total variance accounted for being 3 for the fourth factor, 2 for the fifth, and 1% for the sixth factor. None of these factors, however, makes any interpretive sense.

Graphic Representation of Factors 1 and 2

A number of scales with high loadings on the first two factors are plotted two-dimensionally in Figure 1, for illustrative purposes. The axes in the Figure are drawn at right angles since the Quartimax rotation leads to an orthogonal solution. The first dimension, that of Professional Evaluation, in a sense sets the image that the raters have of "good" professional behavior; that is, responses involving connotations of acceptance, sensitiveness, relaxedness, muteness, calmness, caution, and so on. The second dimension, Precision/Potency, is clearly more allied, as noted earlier, to the original concept of Activity, and its assumed attributes of Lead and Ambiguity.

Construction of the Three-Dimensional Model

A model representing the positions of the 10 therapist responses in "semantic space"

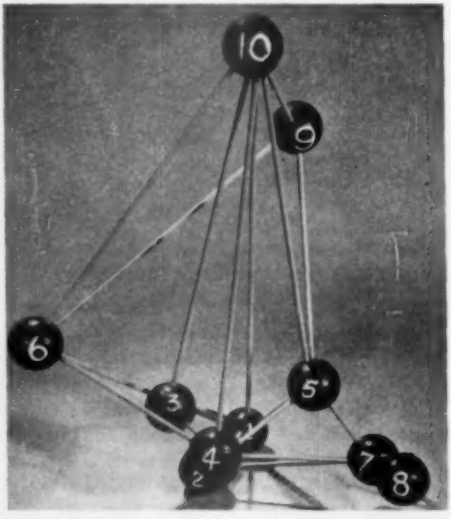

FIG. 2. Three-dimensional model of the 10 therapist responses in "semantic" space.

was constructed[3] from the *Generalized Distance Formula* (Osgood et al., 1957, pp. 90–97). For this purpose the raw scores on three pairs of scales were used to compute values of *D*. The three pairs of scales were selected on the basis of their having high, significant, and relatively pure loadings, respectively, on Factors 1, 2, and 3. The scales were: foolish-wise and valuable-worthless on Factor 1, colorless-colorful and energetic-inert on Factor 2, and inferential-logical and intuitive-rational on Factor 3 (see Table 3). Following Osgood's procedure, the model was represented by 10 rubber balls 1 inch in diameter, fitted together in three-dimensional space with aluminum rods of $\frac{1}{8}$ inch gauge, cut to appropriate lengths. The model is presented in Figure 2. Slight but not at all serious difficulty was experienced in fitting all of the distances exactly, thus suggesting that three dimensions account quite well for the obtained data.

Table 1 earlier presented the 10 verbal responses, and a rationale for their inclusion in the study was given. Figure 2 illustrates that, by and large, a priori expectations are borne out; and the model thus has considerable face validity. It will be recalled that Response 1 consists of "Hm-hm," while Response 10 smacks of a rare attempt to persuade the patient. These two responses are the most distant from each other in the model. Responses 2 and 3 are similar, low active facilitation responses, and they are appropriately close to each other in Figure 2. Responses 4 and 5 refer to more sharply (but equally) focused facilitating responses, and they also are quite close to each other in the model. Response 6 refers to a highly specific, objective (and presumedly uncharged) question; it is unique in the set of 10, and accordingly rather solitary. Responses 7 and 8, on the other hand, constitute interpretive operations of about equally moderate depth, and are adjacent in Figure 2. The ninth response refers to a supportive/reassurance operation, clearly discriminated in the three-dimensional model from other responses. The relative positions of the 10 balls in the model thus accord largely with purely subjective clinical "feel," and agree very well with empirically observed groupings of the responses found in the earlier one-dimensional study.

[3] Thanks are acknowledged to Michael S. Black, now at the University of Illinois, for painstakingly constructing this model.

Discussion

The general findings are reminiscent of two earlier studies. One, published by Fisher (1956), concerned ratings of "plausibility" versus "depth" of interpretive responses. Fisher showed strong, significant relationships between ratings of depth of therapist interpretive operations, and ratings obtained when the working dimension was Plausibility. The findings of the present study independently suggest Fisher's results to be extremely plausible! Presumably there is a (theoretically) infinite number of "one-dimensional" scale labels that would give approximately comparable results in any such rating study.[4] This consideration argues, of course, for the hypothesis that regardless of what *instruction* one gives to a rater, he will, in the final analysis, rate according to certain internal mediating cues which only partly correspond with whatever *explicit* cues the experimenter is trying to communicate. Fisher has drawn attention to the essence of the problem here involved. The present findings rather forcefully suggest that upon closer inspection, some of the dimensions of therapist verbal behavior frequently studied empirically may turn out, operationally speaking, to be one and the same.

A second finding of which the present ones are reminiscent was published by Raush et al. (1956). While those authors concluded that on the whole they could not find evidence of

[4] As a matter of fact, a set of 35 abstract descriptions of therapist verbal responses were recently sorted here by completely naive, unsophisticated, freshman undergraduates, along "any increasing dimension that you think would be appropriate." The rank orderings of these therapist responses sorted by 17 subjects showed a rho of .87 with ranked mean ratings (of the same 35 responses) made by 20 Board-certified psychiatrists given a working definition of Activity Level in terms of Ambiguity, Lead, and Inference, and asked to sort on this dimension! These data are unpublished. They lead one to the opinion that there is a rather basic cultural uniformity in discriminatory reactivity to verbal statements, probably because verbal statements both define and reflect the fundaments of a relationship between two people.

multidimensionality in ratings of their Depth of Interpretation data, one of their studies, nevertheless, did yield three-dimensionality. The first dimension was clearly one of depth; the second was called Ambiguity; while the third was not identified. The authors later concluded that the second dimension was specific to the raters and materials employed. While the adjectival scale deep-shallow used in the present study is in no real sense equivalent to depth as defined by Raush et al., it is nonetheless interesting to note that such a scale is clearly more highly loaded on the Professional Evaluation factor than on the Precision/Potency factor (see Table 3, Line 37). Were there any relationship between the deep-shallow scale and the defined concept of depth, then such would support the hypothesis that under at least some conditions the depth dimension is perhaps one of evaluation, rather than one of ambiguity or precision. These similarities are, of course, only peripheral. It is also to be noted that Osburn (1951) found no evidence of multidimensionality among his ratings of Ambiguity.

The results confirm the prior impression that raters do indeed tend to react to statements in the semantic differential rating situation with an attitude which is primarily evaluative, and that only in the second place, as it were, do they concern themselves with the degree of ambiguity, clarity, activity, precision, and focus of a response. In complementary fashion, it is a comforting finding that judgments of the second kind are not, after all, necessarily tinged with evaluative considerations. Presumably in some experimental rating situations attributes of evaluation and precision may be confounded and thus heighten error, if the subject is forced arbitrarily to rate along an inadequately defined scale. But happily, the two influences in the present study appear to be quite independent.

Finally, it should fairly be said that Osgood and his colleagues have clearly documented the consistent, primary emergence of an evaluative factor in widely differing contexts—even in the ratings of sonar signals. It is thus in one respect not at all surprising that a similar finding was made in the study reported here, and that a fusion of his activity and potency factors emerged second. It remains to be seen whether the 10 *therapist responses themselves* largely hold up to a one-, two-, or three-dimensional hypothesis, when rated by three independent groups of subjects upon each of the three dimensions uncovered. Such a study is planned.

Summary

Thirty-five Board-certified psychiatrists rated 10 bona fide therapist verbal responses against 40, seven-point, bipolar adjectival scales, chosen to correspond with the hypothetical variables of Ambiguity, Lead, Inference, Stressfulness, Evaluation, Potency, and Activity. The matrix of intercorrelations among the 40 scales was analyzed by Thurstone's complete centroid method, and the centroid was rotated via the quartimax method, maintaining orthogonality. The first factor, accounting for 33% of the total variance, was one of Professional Evaluation; the second, accounting for 18% of the total variance, was one of Precision/Potency. The third factor discussed (accounting for only 4%) was one of Subjectivity/Objectivity. The results thus indicate that while ratings were made primarily on an evaluation dimension and secondarily on a dimension of precision and potency (ambiguity), the two dimensions are independent. Results are discussed from the standpoints of their limitations and generalizability, the need for further experimentation, and the findings of other investigators.

REFERENCES

Bordin, E. S., Cutler, R. L., Dittmann, A. T., Harway, N. I., Raush, H. L., & Rigler, D. Measurement problems in process research in psychotherapy. *J. consult. Psychol.*, 1954, **18**, 79–82.

Coombs, C. H. Mathematical models in psychological scaling. *J. Amer. Statist. Ass.*, 1951, **46**, 480–489.

Deutsch, F., & Murphy, W. F. *The clinical interview.* Vol. 2. New York: International Univer. Press, 1955.

Finesinger, J. E. Psychiatric interviewing: Principles and procedure in insight therapy. *Amer. J. Psychiat.*, 1948, **105**, 187–195.

Fisher, S. Plausibility and depth of interpretation. *J. consult. Psychol.*, 1956, **20**, 249–256.

Fruchter, B. *Introduction to factor analysis.* New York: Van Nostrand, 1954.

Harway, N. I., Dittmann, A. T., Raush, H. L., Bordin, E. S., & Rigler, D. The measurement of

depth of interpretation. *J. consult. Psychol.*, 1955, **19**, 247–253.

HOWE, E. S., & POPE, B. An empirical scale of therapist verbal activity in the initial interview. *J. consult. Psychol.*, 1961, in press.

NEUHAUS, J. O., & WRIGLEY, C. F. The quartimax method: An analytic approach to orthogonal simple structure. *Brit. J. statist. Psychol.*, 1954, **7**, 81–91.

OSBURN, H. G. An investigation of the ambiguity dimension of counselor behavior. Unpublished doctoral dissertation, University of Michigan, 1951.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: Univer. Illinois Press, 1957.

POPE, B., HOWE, E. S., & FINESINGER, J. E. Quantitative studies of therapist verbal activity. Paper presented at American Psychological Association, Cincinnati, September 1959.

RAUSH, H. D., SPERBER, Z., RIGLER, D., WILLIAMS, JOAN V., HARWAY, N. I., BORDIN, E. S., DITTMANN, A. T., & HAYS, W. A dimensional analysis of depth of interpretation. *J. consult. Psychol.*, 1956, **20**, 43–48.

SNYDER, W. U. (Ed.) *Group report of a program of research in psychotherapy*. University Park: Pennsylvania State Univer., 1953.

SPEISMAN, J. C. Depth of interpretation and verbal resistance in psychotherapy. *J. consult. Psychol.*, 1959, **23**, 93–99.

THURSTONE, L. L. *Multiple factor analysis*. Chicago: Univer. Chicago Press, 1947.

(Received June 13, 1960)

Journal of Consulting Psychology
1961, Vol. 25, No. 4, 304–311

NEED VALUE AND EXPECTANCY INTERRELATIONS AS ASSESSED FROM MOTIVATIONAL PATTERNS OF PARENTS AND THEIR CHILDREN[1]

FORREST B. TYLER, BONNIE B. TYLER, AND JANET E. RAFFERTY
Southern Illinois University

The nature of the relation between need value (goal value) and expectancy has been approached in two somewhat different ways. In more controlled laboratory studies (Crandall, Solomon, & Kellaway, 1955; Edwards, 1955; Feather, 1959; Irwin, 1953; Marks, 1951; Worell, 1956), the subject's preference for known alternatives or the subject's willingness to bet on an outcome with a controlled actuarial probability of occurrence has been used as a measure from which expectancy, need value, and the relation between them have been inferred. In more broadly conceived research such as that of Atkinson and his colleagues (1958), story themes have been used to study strength of a single motive such as need for achievement. Results from both of these types of studies have demonstrated the importance of the two concepts of need value and expectancy in dealing with a variety of problems ranging from predictions of animal behavior in choice situations to analysis of thematic test responses in personality research (Feather, 1959).

Among theoretical approaches to the study of need value and expectancy there have been differences as to the relation between these two variables. Rotter (1954) considers need value and expectancy to be independent, as does Edwards (1955) in his *SEU* model. In contrast, Atkinson (1958) and others (Feather, 1959) consider them interrelated. Empirical evidence indicates that they are statistically interrelated under some conditions, but that the relation found varies for different probability levels (Crandall et al., 1955) and for different goal value levels (Worell, 1956). Further, any evident relation decreases when a "premium" (i.e., a high goal value) is placed on accuracy of estimates of occurrence (Crandall et al., 1955; Worell, 1956). Also, the relation may be less direct in adults than in children (Crandall et al., 1955; Irwin, 1953; Marks, 1951). Finally, the nature of the relation may be different for achievement and nonachievement situations (Atkinson, 1958; Feather, 1959; Marks, 1951; Worell, 1956).

The objective of the present paper is to present findings pertinent to three questions concerning need value–expectancy relations that are of theoretical as well as practical importance. These questions concern: the independence of need value and expectancy, the relation between need value and expectancy in parents in contrast to that relation in children, and the relation between need value and expectancy for recognition-status motives in contrast to other motives.

Methodological Procedures and Rationale

Subjects

The subjects were three samples of children, aged 2–6 to 5–0, enrolled in a cooperative preschool located on a college campus, and the parents of those children. The three samples included, respectively, 18, 16, and 11 children; total *N*s were 45 children (20 boys, 25 girls), 45 fathers, and 45 mothers. Forty-three of the 45 sets of parents were college student or faculty families; the remaining 2 families were from the local community.

Collection of Data

Parent data were collected in a 100-question free-response interview structured to elicit information

[1] This investigation was supported by a research grant, M-1137, from The National Institute of Mental Health, United States Public Health Service and by research funds from Southern Illinois University.

concerning parental child-rearing motivations. Each interview required approximately 2 hours to complete and was tape recorded. Interviews were conducted by two trained interviewers, each of whom interviewed half the fathers and half the mothers, but never both members of the same family. Child data were obtained by two trained observers who made narrative records of each child's behavior for 5-minute periods during regular preschool activities. Each child was observed approximately 300 minutes during the first 3 months of school in the fall, with observations for each child distributed equally over the 3-month period. The children were observed in a systematic predetermined manner designed to prevent selective biases. All parent interviews and child observations were typed and data analyses were made from typescripts.

Rating Procedures

The present study has involved categorizing and quantifying parent interview responses and child preschool behaviors on the same motivational variables, while at the same time avoiding contamination in the assessment of parental and child protocols. Accordingly, two research teams operated independently—one working with the child data and one with the parent data—to develop explicit operational definitions for all concepts, and to construct scoring-by-example manuals for the parent data and for the child data using these definitions.

The task of constructing parallel operational definitions of need value and expectancy for parents and for children's behavior was accomplished as follows. All records were analyzed by referents, or behavioral units. Each unit contained information concerning: (*a*) the stimulus context at the moment (interview question for parents; nursery setting or specific behavioral stimuli for children); (*b*) the goal-direction of the response (for the parent this goal-direction was inferred from content of verbal statements; for the child it was inferred from the nature of his behavior, e.g., friendly behaviors are considered to be love and affection goal-directed); and (*c*) other characteristics of the response (e.g., for the parent, persistence, statement of anticipation or dread as accompanying a behavior, etc.; for the child, persistence, constructive-nonconstructive or defensive nature, accompanying verbalization, etc.).

The motivational categories used for classifying these data with regard to goal-directional characteristics are derived from Rotter (1954). As utilized in the present study, they have been given the following operational definitions:

Recognition-Status (R-S): Parents—concern with child's achievements, teaching skills to child, being known as a good parent, doing what a parent should. Children—calling attention to achievement, behaviors, attributes, or possessions; conformity to or imitation of achievement or other socially approved behaviors; attention-getting activities.

Love and Affection (L&A): Parents—concern with play or companionship with child, interest in child's happiness. Children—cooperative play, friendly behaviors, sharing, helping, sympathetic behavior, seeking affection.

Dominance (Dom): Parents—concern with controlling child, teaching obedience, molding child. Children—commanding others, making demands, aggression, controlling others' activities.

Protection-Dependency (P-D): Parents—(not relevant to child-rearing motivations of parents with preschool children). Children—seeking help, information, permission, comfort or consolation, and intervention of others to prevent frustration.

Independence (Ind): Parents—(not relevant, since independence satisfactions by definition are self-mediated). Children—individualized activity, self care.

Rationale

The following general rationale was developed as a basis for differentiating expectancy and need value referents for each motivational category.

Need value. In general, there is agreement among psychologists that there is a direct relationship between the value of a goal or reinforcement and choice preference for that goal. Accordingly, as a basis for inferring need value (NV) ratings on the seven-point scale used, the following operations were set up.

1. Strength of NV rating is determined by:

a. Variations in stimulus cues as follows: For both parents and children the fewer the stimulus cues, or the less the "stimulus pull," the higher the need value rating given. So if there are very few stimulus cues present—as when an interview question is directed toward eliciting a response in one need area and the response given is directed toward another need—the response is given a relatively high NV rating.

b. Variations in response characteristics as follows: For children, persistence of response leads to a higher rating. For parents, statements concerning persistence, or statements concerning instigation of the indicated kind of activity in situations with few cues present lead to a higher rating.

2. Need category to be scored for NV is indicated as follows:

a. Goal direction [2] of response, either interview statement or observed behavior, is scored.

b. Goal direction indicated by the situational structure is scored if the situation is "maximized" for response in that direction. This criterion leads to scoring for nonoccurrence of a response when a maximally structured situation is rejected completely by the person responding; this rejection is considered evidence of very low NV in that area. The response direction is also scored as indicated earlier and given a high NV rating if it occurred when there were very few stimulus cues present to elicit it.

This procedure of scoring NV in an inverse relation to the strength of eliciting stimuli is roughly comparable to procedures followed in projective test-

[2] For a more detailed account of determination of goal-direction of responses, see Tyler (1960).

ing in which attempts have been made systematically to reduce clarity of stimulus cues as a basis for getting at important personality dimensions, and in which unusual or atypical responses are considered most significant (Lindzey, 1952). Scoring nonoccurrence of responses for low NV under specified conditions is less generally accepted, but seemed justifiable and consistent with the broader rationale on which NV strength was determined. The behavioral variable used, persistence, is traditionally considered an indication of the strength of motivation to achieve a goal, and consequently of the value of the goal to the person behaving.

Expectancy. There is considerably less agreement among psychologists as to the relationship between expectancy (Ex) or subjective probability and choice preference for a goal. One point of view is that there is a direct relation between these two (Atkinson, 1958). A somewhat conflicting view is that the effect of low Ex of goal attainment is to lead to the occurrence of "defensive" behaviors with regard to important goals (Eriksen, 1950; Rotter, 1954). Much attention has been given in personality research to study of defenses, even to the extent at times of assuming that nonoccurrence of pertinent behavior must be defensiveness (Zuk, 1956). This extreme position is not taken here. The position is taken that defensive behaviors are not categorically different from other behaviors; rather they indicate the low end of a continuum of constructiveness of goal directed activity. That is, the constructiveness of behaviors is considered to reflect directly the subject's level of subjective probability. Consequently the following measures of expectancy were set up.

1. Need category to be scored for Ex is indicated by goal direction of response.

2. Strength of Ex rating is indicated by:

a. Variations in response characteristics as follows: For children, ratings range from lowest for withdrawal behaviors, clearly defensive behaviors (e.g., disruptive or nonsocially approved behaviors), and tentative abortive behavior attempts; up the scale to highest ratings for direct, socially approved interactions which include indicators of high confidence (e.g., direct smiling approaches to new children, etc.). For parents, ratings range from lowest for statements of dismay, loss of self-control, and failure and frustration over attempts to attain goals; up the scale to highest ratings for statements of anticipation of interacting with children, joy at interactions together, etc.

b. Variations in stimulus cues also affect Ex as follows: For children, since data used were records of direct behavior in relatively free situations (i.e., no "forced choices" of a controlled sort existed as they might in experimental situations), it was thought that the levels of expectancy which were quite high might not be differentiable except in situations where a few stimulus cues were present. Consequently it was decided to give the highest ratings only to high confidence responses which occurred in situations where there were a minimum of relevant stimulus cues. Although this method of scoring child Ex makes possible a slight overlap between child NV and Ex scores, it was deemed essential for the reasons indicated. For parents it was possible to structure questions to elicit expectancy statements independent of need value, so no comparable variations in scoring were required for parental expectancy measures.

Correlations reported are Pearson product-moment r's (McNemar, 1955) based on mean scores computed from all the ratings given to an individual. Differences between correlations are assessed using Fisher's z' transformation (McNemar, 1955).

Findings

This study has been concerned with charting relationships among motivational characteristics in parents and in children. For that reason findings obtained are meaningful for testing hypotheses primarily in a construct validity fashion (Cronbach & Meehl, 1955). Correlations reported are examined to determine whether they are consistent with relationships hypothesized to exist from specified theoretical points of view. Their support for any such aspects of a "nomological net" constitutes construct validation.

Interrater reliabilities for the parent and child scoring manuals are reported in detail elsewhere (Rafferty, Tyler, & Tyler, 1960; Tyler, Tyler, & Rafferty, 1959). Those figures can be summarized by noting that median interrater reliabilities for these manuals range from .68 to .84 for samples of 18 and 16 subjects.

Correlations reported will be for the total 45 subjects from all three samples studied. The indicated analyses have been run for the separate samples, but the variability in findings between samples is within the limits expected by chance. Hence, they are not reported here.

The need categories used in this study are Recognition-Status (R-S), Love & Affection (L&A), Dominance (Dom), Protection-Dependency (P-D), and Independence (Ind). For each category, both Need Value (NV) and Expectancy (Ex) scores have been obtained.

It was necessary to assess the independence of the need categories to determine whether separate NV–Ex interrelationship analyses were justifiable. The pertinent interneed correlations among NV measures can be summarized briefly. For fathers, mothers, and

children the R-S and L&A NVs are clearly independent, and there is a moderate positive relationship ($r = .40+$) between R-S and Dom NVs in all three sets of measures. A somewhat smaller but statistically significant inverse relationship ($r = -.30+$) exists between L&A and Dom NV measures for parents. For girls, this L&A–Dom NV relationship also tends to be negative, though for boys it is positive. However, neither of these correlations approaches statistical significance. Child interneed comparisons do yield marginally significant (.05 level) r's between R-S and P-D ($r = .34$) and between L&A and Ind ($r = -.31$). Nevertheless, 7 of the 10 interneed correlations for all children are nonsignificant; 8 of the 10 for girls alone are nonsignificant; and all 10 are nonsignificant for boys. In general it seems appropriate to conclude that these correlations indicate sufficient overall independence among these NV measures to warrant consideration of them as functionally distinct motivational categories.

When interneed correlations among expectancy measures are examined more indication of a consistent pattern of positive interrelationships is found than was the case among need value measures. However, this pattern seems to be primarily a function of a generalized level of Ex among the R-S, L&A, and Dom needs which holds for mothers and girls. This generalized relationship is moderate ($r = .40+$) for mothers, and moderate to high (r's range from .55 to .75) for girls. In contrast, for fathers the only interneed Ex relationship is a moderate one ($r = .45$, significant at .01 level) between L&A and Dom, while for sons these three Ex measures are not significantly interrelated. The nature of the sex differential for this generalized Ex pattern is indicated even more clearly by the fact that boy-girl differences are significant (two at the .10 level, one L&A–Dom at the .05) for these three comparisons, and father-girl differences are significant on two of them (R-S–L&A r's at .05 level, L&A–Dom r's at .01 level). The differences between mothers interneed Ex correlations and those of fathers and boys also tend to indicate a more generalized Ex for mothers, but none of the correlational differences is significant between mothers and sons, and only one (L&A–Dom, D significant at .05) is significant between mothers and fathers. Even though this somewhat generalized Ex for females is noted, it should also be pointed out that 6 of the 10 interneed expectancy r's for girls are not significantly interrelated; nor are the 3 Ex measures for mothers so interrelated as to warrant considering them as 3 measures of the same variable. For fathers, only 1 of the 3 comparisons yields even a moderate r, and for boys only 1 of 10 (Dom–Ind $r = .46$, significant at .05 level) achieves significance. Thus there seems to be sufficient independence among these expectancy variables to justify consideration of them as operationally independent.

Results of analyses of NV–Ex interrelationships are reported in Table 1. Findings pertinent to answering the first general theoretical question of the independence of NV and Ex can be summarized as follows:

1. Fathers' NV and Ex scores show a slight

TABLE 1

INTRANEED CORRELATIONS BETWEEN NEED VALUE AND EXPECTANCY MEASURES

Need Category	Parents		Children		
	Fathers ($N = 45$)	Mothers ($N = 45$)	Boys ($N = 20$)	Girls ($N = 25$)	Total ($N = 45$)
Recognition-Status	.19	.00	.23	.60	.40**
Love and Affection	.27	.41**	.22	−.13	.00
Dominance	.21	.07	.71**	.79**	.76**
Protection-Dependency			.66**	.62**	.64**
Independence			.82**	.62**	.72**

Note.—Level of significance: $N = 45\ 25\ 20$
* $r_{.05} = .29\ .40\ .44$
** $r_{.01} = .40\ .51\ .56$

TABLE 2

DIFFERENCES BETWEEN NEED VALUE–EXPECTANCY INTERCORRELATIONS AMONG SUBJECT GROUPS

Need Category	Fa-Mo	Fa-Child	Fa-Boy	Fa-Girl	Mo-Child	Mo-Boy	Mo-Girl	Boy-Girl
Recognition-Status NV–Expectancy	.192	.232	.042	.501	.424	.234	.693**	.459
Love and Affection NV–Expectancy	.159	.277	.053	.408	.436*	.212	.567*	.355
Dominance NV–Expectancy	.144	.782**	.673*	.857**	.926**	.817**	1.001**	.840
Protection-Dependency NV–Expectancy								.068
Independence NV–Expectancy								.432

Note.—Fathers (N =45), Mothers (N =45), Boys (N =20), and Girls (N =25).
* Significant at or beyond the .05 level.
** Significant at or beyond the .01 level.

positive interrelation of the order of .20+, but they are not sufficiently interrelated for that fact to be significant statistically.

2. For mothers, any NV–Ex relation is confined to a moderate one ($r = .41$, significant at .01 level) within the L&A need category.

3. For children, there is a high positive intercorrelation in three need categories, Dom, P-D, and Ind. There is also a high NV–Ex r (.60, significant at .01 level) for girls on the R-S variable. However, NV and Ex measures for children on L&A are completely independent. It would seem that the NV–Ex interrelations obtained are relatively specific, and are a function of at least the following factors: (*a*) age of subject, e.g., parents show a closer NV–Ex relation on L&A than do children; (*b*) sex of subject, e.g., boys and girls differ in degree of NV–Ex relationship on R-S; and (*c*) motivational category under consideration, e.g., boys and girls have a quite different NV–Ex relationship on L&A than on Dom.

The second question of whether NV and Ex are less independent in children than in adults can be answered by reference to Tables 1 and 2. For girls, it can be seen that their need value–expectancy scores are significantly more closely intercorrelated for R-S motives (girl-mother difference significant at .01, girl-father at .06) and for Dom motives (both comparisons significant at .01 level) than is the case for their parents. However, the converse is true for L&A motives since the NV–Ex r for mothers is significantly greater (.05 level) than that for daughters, and the comparable r for fathers is greater than that for daughters, though this latter difference does not reach an acceptable level of statistical significance. For boys, a similar pattern exists though the differences are not as marked. The greater NV–Ex interrelation on R-S and Dom holds for boys in relation to parents, but is statistically significant only for the latter. Also, the greater NV–Ex interrelation for parents on the L&A category holds with respect to boys and parents, but it does not achieve significance.

These comparisons of parental and child NV–Ex relations do not indicate consistently closer NV–Ex relations for children; rather, they seem to indicate that whether parents or children will be found to have a closer NV–Ex relation is a function of the motivational category under consideration.

The third question asked is that of the relation between NV and Ex for R-S motives in contrast to other motives. The hypothesis for which this comparison is pertinent is one advanced by Atkinson (1958) which states that there is an inverse relationship between incentive value and subjective probability for need Achievement which does not maintain

for other needs. Data to test this hypothesis are the relationships presented in Table 1, with the pertinent comparisons being those between rows for each subject group. For fathers there are no differences from need to need in the NV–Ex relation. For mothers, the R-S and Dom NV–Ex relations are not different, though the comparable L&A relation is significantly greater than comparable comparisons on the R-S variable ($D = .436$, significant at .05 level), and the Dom variable ($D = .366$, significant at .10 level). For children the NV–Ex relation on R-S is intermediate in relative as well as absolute magnitude. It is greater ($D = .424$, significant at .06 level) than the comparable relation for L&A and less than that on the other three need variables (P-D $D = .334$, nonsignificant; Ind $D = .484$, significant at .05 level; Dom $D = .572$, significant at .01 level). The findings are not consistent with the hypothesis advanced by Atkinson since for none of the three groups is there clearly a less direct or more inverse relationship between NV and Ex for R-S motives than for all other motives.

Discussion

The three theoretical questions concerning need value–expectancy interrelationships which are the primary focus in this article concern (*a*) the independence of need value and expectancy, (*b*) the greater independence of need values and expectancies in adults than in children, and (*c*) the less direct relationship between need value and expectancy variables for recognition-status motives than for other motivational categories. It should be noted that previous data from which answers to these questions have been inferred have frequently been derived in controlled experimental situations where known expectancies and goal preferences could be systematically manipulated. In contrast, the findings of this study are based on more molar data with mean scores for each subject derived from ratings of a number of referents (answers to interview questions, or observed behaviors). Although it is not appropriate to assess from the present findings the validity of results obtained in specific experimental settings, it is possible to determine from these results the representativeness of those more molecular situations and the generality of conclusions derived from them.

The question of a general need value–expectancy relation which holds for all needs and all subjects has been raised by Rotter (1954), Atkinson (1958), Edwards (1955), and others (Feather, 1959). Neither a general position of NV–Ex independence (Edwards, 1955; Rotter, 1954) nor a position of interdependence (Atkinson, 1958) can be supported by the findings of this study. Rather, it would seem that need value–expectancy interrelations are specific to the sex and age of the subjects studied, and to the motivational category measured. Although fathers yield a stable nonsignificant NV–Ex relationship for all need comparisons, none of the other subject groups do. Further, the NV–Ex relationship for parents is different from that for children on both L&A and Dom, and this interrelation for girls on R-S is significantly different from that for boys, for fathers, and for mothers on the R-S variable. In general, these parent-child differences in need value–expectancy relationship such as that obtained on the L&A variable support the conclusion that adult motivational patterns are not clearly established during the preschool years. Further, the obtained sex differences, such as that between boys and girls on R-S, provide substantial evidence of sex-linking in patterns of motivational development. The specificity of relations found leads to underscoring the need for caution in generalizing from the results of one study to motivational characteristics of the population at large.

The second question which concerns a closer relationship hypothesized between need values and expectancies in children than in parents is derived from a suggested explanation for discrepancies in results obtained by Crandall et al. (1955) and Irwin (1953) in contrast to those of Marks (1951). Crandall has indicated that the smaller effect of reinforcement values on expectancy statements found in his work and that of Irwin, in comparison to Mark's findings, may be a function of greater susceptibility to reinforcement value effects in younger children, since Marks used 9 to 11 year olds as subjects, and Crandall and Irwin used college students. If such a phenomenon does exist, then preschool children's NV and

Ex measures should be more closely related than is the case for adults (i.e., their parents). The results obtained suggest that this explanation is an oversimplification of the relevant variables. Specifically, on the Dom variable the children show a greater NV–Ex interrelationship than the parents do, on the L&A variable the parents show a greater relationship than the children, and on the R-S variable the boys show a similar NV–Ex relationship to that of their parents, although the girls show a closer relationship.

The third theoretical question concerns the hypothesis advanced by Atkinson (1958) that there is an inverse relationship between incentive value and subjective probability for n Ach which does not maintain for other needs. As noted earlier there is not complete overlap between the R-S motive category and the n Ach category, but R-S would seem to be subsumed within the latter category. Consequently, findings concerning it should be pertinent. Findings from the present study show that the indicated R-S NV–Ex correlation is not significantly closer to an inverse relationship for R-S than for all other needs for any of the four groups (fathers, mothers, boys, girls) analyzed. In fact, for each group of subjects there is at least one other variable on which the NV–Ex relationship is no different from or less direct than the R-S NV–Ex relationship. Therefore, although these findings support an assumption of specificity of NV–Ex interrelationships to age and sex of subjects, and to motivational category under study, they are not consistent with Atkinson's hypothesis. It may be that this discrepancy is a function of the lack of comparability between the more molar level of measurement in the present study and the more specific and controlled measurement situations in which his work has been conducted. This discrepancy may also suggest that the inverse relation between subjective probability and incentive value which seems to hold in his experimental situations may not be an important determinant of behavioral choices of the sort on which the findings of the present study are based. However, the obtained intersubject differences on the R-S variable warrant particular consideration, since many investigations of motivation have focussed on achievement-type situations and tasks. Although the recognition-status motive includes only other-mediated achievement goals, so it is not completely equivalent to n Ach, it does fall within the achievement category. To the extent that these concepts overlap, the present findings indicate clearly that for preschool girls importance of achievement goals and subjective probability of attaining those goals are much more closely linked than is the case for preschool boys or for parents. Since this differential does not hold for the mothers' child-rearing motivations, it would seem that girls may acquire this distinction at a later date than is the case for boys. Nevertheless, these findings tend to confirm the results of Veroff (1953), which indicate that there are significant sex differences in response to achievement situations.

Summary

In the present paper relations among the need values (NVs) and expectancies (Ex's) of child-rearing motivations of parents and of motivations of their preschool children are analyzed. Subjects were 45 families with children enrolled in a cooperative preschool in a university housing project. Parental motives for recognition-status (R-S), love and affection (L&A), and dominance (Dom) were assessed from free-response interviews; child motivations for the same three motives, and for protection-dependency (P-D), and independence (Ind) were assessed from narrative observations of preschool activities. Adequate interrater reliabilities for both parent and child score-by-example manuals for rating the indicated protocols are reported.

Patterns of intercorrelation among the motivational variables measured support the conclusion that they are operationally independent. Need value–expectancy correlations obtained are viewed as evidence in a construct-validation sense of the validity of certain hypotheses which have been advanced by others concerning theoretical interrelations between need value and expectancies. On the basis of these findings, there is no support for the hypothesis that there is either a general pattern of independence or a general pattern of interrelation between NVs and Ex's which holds for all subjects and all motivational cate-

gories. Rather, there is support for the position that NV and Ex interrelations are specific to the need category under study, and to age and sex of subjects. The hypothesis that NVs and Ex's are less independent in children than in adults is likewise not supported generally, since this interrelation also seems to be a function of the motivational category under study, with parents differentiating more clearly on one of the three common motivational categories, and children differentiating more clearly on one. The hypothesis that there tends to be an inverse relation between NVs and Ex's for R-S motives which does not maintain for other motives could not be supported. Although there were clear NV–Ex correlation differences from need category to need category, for none of the subject groups was the R-S NV–Ex *r* clearly less direct than was the case for the other motives.

REFERENCES

ATKINSON, J. W. (Ed.) *Motives in fantasy, action, and society*. Princeton, N. J.: Van Nostrand, 1958.

CRANDALL, V. J., SOLOMON, D., & KELLAWAY, R. Expectancy statements and decision times as functions of objective probabilities and reinforcement values. *J. Pers.*, 1955, **24**, 192–203.

CRONBACH, L., & MEEHL, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, **52**, 281–302.

EDWARDS, W. The prediction of decisions among bets. *J. exp. Psychol.*, 1955, **50**, 201–214.

ERIKSEN, C. W. Some implications for TAT interpretation arising from need and perception experiments. *J. Pers.*, 1950, **19**, 282–288.

FEATHER, N. T. Subjective probability and decision under uncertainty. *Psychol. Rev.*, 1959, **66**, 150–164.

IRWIN, F. W. Stated expectations as functions of probability and desirability of outcomes. *J. Pers.*, 1953, **21**, 329–335.

LINDZEY, G. Thematic Apperception Test: Interpretive assumptions and related empirical evidence. *Psychol. Bull.*, 1952, **49**, 1–25.

MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1955.

MARKS, ROSE W. The effect of probability, desirability, and "privilege" on the stated expectations of children. *J. Pers.*, 1951, **19**, 332–351.

RAFFERTY, JANET E., TYLER, BONNIE B., & TYLER, F. B. Observations of free play situations as a method of personality assessment. *Child Develpm.*, 1960, **31**, 691–702.

ROTTER, J. B. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954.

TYLER, BONNIE B., TYLER, F. B., & RAFFERTY, JANET E. A systematic approach to interviewing as a method of personality assessment. Paper read at American Psychological Association, Cincinnati, September 1959.

TYLER, F. B. A conceptual model for assessing parent-child motivations. *Child Develpm.*, 1960, **31**, 807–815.

VEROFF, J., WILCOX, S., & ATKINSON, J. W. The achievement motive in high school and college age women. *J. abnorm. soc. Psychol.*, 1953, **48**, 108–119.

WORELL, L. The effect of goal value upon expectancy. *J. abnorm. soc. Psychol.*, 1956, **53**, 48–53.

ZUK, G. H. The influence of social context on impulse and control tendencies in preadolescents. *Genet. psychol. Monogr.*, 1956, **54**, 117–166.

(Received June 13, 1960)

Journal of Consulting Psychology
1961, Vol. 25, No. 4, 312–315

CROSS-VALIDATION OF A RORSCHACH CHECKLIST ASSOCIATED WITH SUICIDAL TENDENCIES

IRVING B. WEINER [1]

University of Rochester School of Medicine and Dentistry

Although numerous methods for assessing suicidal potential by means of the Rorschach test have appeared in the literature, failure to cross-validate "signs" and "configurations" supposedly indicative of suicidal tendencies has cast doubt on the utility of these methods (Fisher, 1951; Sakheim, 1954). However, recent work by Daston and Sakheim (1960) using Martin's checklist, 17 signs empirically derived from comparisons between suicidal and nonsuicidal psychiatric patients, has yielded promising results. Daston and Sakheim compared the Rorschach protocols of patients who were nonsuicidal with those of patients who had attempted suicide or had actually taken their own lives. In their study there were virtually no differences between completed suicide and suicide attempt groups on Martin's checklist, but both of these groups received significantly more of Martin's signs than the nonsuicidal group. Furthermore, 83% of the successful suicides and 72% of the attempted suicides received six or more of Martin's signs, while only 17% of the controls displayed this many signs.

Since the above results were derived from a population of hospitalized male veterans undifferentiated with respect to psychiatric diagnosis, question may be raised concerning the concurrent validity of Martin's checklist for a population less homogeneous than the veteran group. This question in turn points to the importance of evaluating the influence of such variables as age, sex, hospitalization, and nature of psychopathology on the checklist scores. Additionally, since most Rorschach scores are significantly related to the total number of responses given (Fiske & Baughman, 1953), it is pertinent to examine the relationship between response total and checklist score.

[1] Appreciation is expressed to Norman I. Harway for suggestions concerning the manuscript.

Procedure

Two samples were used in this study. The first, which was intended to provide general information about Martin's signs, consisted of all adult patients in a 6-month period from the psychiatric services of a general hospital for whom scorable Rorschach protocols were available. Patients whose primary diagnosis was organic rather than functional in nature were excluded. The sample contained 28 males and 43 females, had a median age of 29.5 years (range 15–55), and was comprised of 42 hospital and 29 clinic patients. The total group is categorized by age, sex, and patient status in Table 1.

TABLE 1

Age, Sex, and Patient Status of Subjects

Age	Number Males		Number Females		Total
	Hospital	Clinic	Hospital	Clinic	
15–21	1	4	6	5	16
21–30	4	5	11	4	24
31–40	6	4	7	4	21
41–55	3	1	4	2	10
Total	14	14	28	15	71

The Rorschach records of these 71 patients, which ranged in number of responses from 7 to 67 with a median of 20.33, were scored for Martin's signs, which are the following: No. $D < 6$ or > 20; $D\% < 60$ or > 79; No. $CF > 0$ to < 3; Total Color $Rs < 1$; *Sum* $C > 1.0$ to < 3.5; C or CF appear first on VIII-X; C or CF with *Sum* $V +$ *Sum* $T < 1$; No. $FV + VF < 1$; *Sum* $V < 1$; *Sum* $V +$ *Sum* $T < 1$; difference between M and *Sum* $C < 1.5$; No. $H + Hd > 6$; No. Categories < 6 or > 13; VIII-X/$R > 29\%$; No. $P < 3$ or > 6; $P < 3$ with $F + \% > 60$; and time first $R < 27$ seconds.

Subsequently the subjects' hospital and clinic records were utilized to assign the subjects to one of three diagnostic categories: *neurosis*, which included cases of conversion hysteria, obsessive-compulsive

neurosis, anxiety reaction, and neurotic depressive reaction; *character disorder,* which applied to instances of personality trait and personality pattern disturbance; and *psychosis,* which included schizophrenic, psychotic depressive, and involutional psychotic reactions. The diagnostic criterion was the label assigned to the patient during that psychiatric contact in which the Rorschach had been given. For hospital patients, these labels were the discharge diagnoses; for clinic patients, the recorded consensus of an intake committee was used. The hospital group contained 13 neurotics, 14 character disorders, and 15 psychotics; for the clinic group, the respective totals were 13, 10, and 6.

Further examination of the records revealed that eight of the subjects had made documented, serious suicide attempts which had precipitated their admission to the hospital. A search of case files for the year previous to the 6-month period of the study yielded an additional sample of 16 patients who had received psychological testing during a hospital admission subsequent to an attempt at suicide. Since this second group had by and large been referred for testing from the same wards by the same personnel and had been evaluated by the same psychologists as the original eight suicide attempts, it was felt appropriate to combine these two groups for purposes of comparison with the 63 nonsuicidal patients originally studied. The suicidal group was composed of 9 males and 15 females, had a median age of 30 years (range 16–55), and had given between 7 and 57 Rorschach responses with a median response total of 18. Seven of these patients had been diagnosed neurotic, nine as character disorders, and eight as psychotic. It may be noted that this rather even nosological distribution of suicidal patients is consistent with the findings of Pokorny (1960). The observed similarity between completed suicide and suicide attempt groups (Daston & Sakheim, 1960) appears to justify noninclusion of actual suicides in this study. The significance of suicide attempts as a precursor of actual suicide, as discussed by Shneidman and Farberow (1957, pp. 3–10) and Robins, Gasner, Kayes, Wilkinson, and Murphy (1959), further supports this decision.

Results

The original 71 subjects received between 4 and 12 of Martin's signs, with a median of 7.29 signs. As the distribution of numbers of signs did not approach normality (Rorschach scores in general are not normally distributed [Fiske & Baughman, 1953]), and as the nature of the sampling did not provide equal numbers of subjects for various subgroupings, the data were treated with nonparametric techniques. The distribution of number of Martin's signs was divided into five approximately equal parts for comparison with the variables of age, sex, patient status, and number of Rorschach responses by means of $5 \times n$ contingency tables. The distributions of age and number of responses were each divided into four equal parts; for the dimension of patient status, hospital (inpatient) and clinic (outpatient) groups were contrasted. None of the obtained χ^2 values for the association between number of Martin's signs and these four variables approached significance.

TABLE 2

Association between Martin's Suicide Signs, Suicide Classification, and Diagnostic Category

Suicide Classification	Diagnostic Category		
	Neurosis	Character Disorder	Psychosis
Suicide attempt:			
Number above Mdn[a]	6	7	6
Number below Mdn[b]	1	2	2
Nonsuicidal:			
Number above Mdn[a]	5	10	11
Number below Mdn[b]	20	9	8

[a] Eight or more signs.
[b] Fewer than eight signs.

In Table 2 are listed the frequencies of suicidal and nonsuicidal patients with different diagnoses who received more or less than the total group median number of Martin's signs. Analysis of these data by Wilson's (1956) method for nonparametric analysis of variance revealed the following: (*a*) number of signs and suicide classification were significantly associated beyond the .005 level of confidence; (*b*) the association between number of signs and diagnostic category was significant beyond the .05 level; and (*c*) there was no significant interaction between suicidal classification and diagnostic category.

The median numbers of signs received by the two groups were 8.83 for the suicide attempts and 7.04 for the controls. A median test indicated that the discrepancy between these values was significant at the .01 level of confidence. No clear cutting score separating suicidal from nonsuicidal patients emerged from the data; the most efficient cutting score, an incidence of eight or more signs, correctly classified 79% of the suicide attempts and 60% of the controls.

The 17 signs were examined individually for their capacity to differentiate between the suicidal and nonsuicidal groups. A series of χ^2

tests revealed that two signs (*C* or *CF* appear first on VIII-X and $P < 3$ with $F + \% > 60$) had been received by significantly ($p < .05$) more of the suicidal than the nonsuicidal group; trends ($p < .10$) in this direction occurred with two signs (No. $CF > 0$ to < 3 and *C* or *CF* with *Sum V* + *Sum T* < 1); one sign (VIII-X/$R > 29\%$) significantly differentiated the two groups in the direction opposite from expectation; and the other signs did not significantly discriminate suicidal from nonsuicidal patients.

Discussion

The data indicate that, within the ranges sampled by this study, scores on Martin's checklist may be interpreted similarly for men and women patients of different ages, whether they are hospitalized or applying for outpatient psychiatric care, and regardless of how many responses they give to the Rorschach. Number of Martin's signs is significantly and positively associated both with likelihood of having made a suicide attempt and with severity of psychopathology, if the labels neurosis, character disorder, and psychosis may be taken as a continuum of increasing emotional disturbance. The significance of both main effects, in the absence of any significant interaction, indicates that although the potential of the checklist to assess severity of pathology may be a topic for further investigation, the capacity of the measure to discriminate suicidal from nonsuicidal patients operates independently of diagnostic category.

The operation of the individual signs merits consideration, although the positive findings do not add much to current conceptions of what kinds of people attempt suicide. The best discriminator between the suicide attempts and the controls, $P < 3$ with $F + \% > 60$, may be interpreted as a rejection of conventional behavior patterns in the presence of adequate reality testing. However, this sign, though of predictive value, is too rare even in suicidal patients to have general descriptive value. The other significant discriminator and the two signs which approach significance (see Results) may reflect difficulty in dealing comfortably with affective stimulation and controlling tendencies toward impulsive behavior. But such speculations do not provide hypotheses about suicidal individuals which have not previously been advanced without recourse to projective test data, and the value of the checklist would seem to lie more in practical application than in contribution to theory. Concerning the several signs which contributed little or nothing to the differentiation between suicidal and nonsuicidal groups, the moral may be iterated that empirically derived items require careful and repeated cross-validation before they achieve status as efficient and reliable predictors.

With regard to future use of Martin's checklist, a final point should be mentioned. Rosen (1954) and Meehl and Rosen (1955) have questioned the efficiency of psychometric instruments for the prediction of rare events such as suicide and argue that the inevitable number of false positives in such endeavors renders even highly valid discriminators impracticable. However, Cureton (1957) has discussed a method of allowing for population base rates in the establishment of cutting scores to predict a dichotomous criterion from a continuous predictor. The necessity is therefore for collection of sufficient normative data to establish the distribution of the predictor, i.e., Martin's checklist, in order that through consideration of suicide rates, efficient prediction of suicidal risks may be implemented. The data of the current study would suggest that the checklist has adequate concurrent validity to justify such future research.

Summary

Rorschach protocols of 71 patients who had been tested during a 6-month period were scored for Martin's suicide signs. The number of signs earned was found to be relatively independent of the age, sex, hospital status, and Rorschach response total of the subjects. Subsequently, the records of 63 of these patients who were nonsuicidal were compared with those of 24 patients who had made suicide attempts. The suicidal group received significantly more of Martin's signs than the controls. A significant positive association was also found between the number of signs received and severity of psychiatric illness. The

absence of any significant interaction between suicidal classification and diagnostic category indicated that the relationship between suicidal disposition and number of signs was similar in the different diagnostic groups, however. The results did not delineate any efficient cutting scores for predictive use. Nevertheless, it is felt that the demonstrated concurrent validity of Martin's checklist justifies further research to develop reliable norms which, when considered in the light of base rate incidence, might facilitate efficient prediction of suicidal tendencies.

REFERENCES

Cureton, E. E. Recipe for a cookbook. *Psychol. Bull.*, 1957, **54**, 494–497.

Daston, P. G., & Sakheim, G. A. Prediction of successful suicide from the Rorschach test using a sign approach. *J. proj. Tech.*, 1960, **24**, 355–361.

Fisher, S. The value of the Rorschach for detecting suicidal trends. *J. proj. Tech.*, 1951, **15**, 250–254.

Fiske, D. W., & Baughman, E. E. Relationships between Rorschach scoring categories and the total number of responses. *J. abnorm. soc. Psychol.*, 1953, **48**, 25–32.

Meehl, P. E., & Rosen, A. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychol. Bull.*, 1955, **52**, 194–216.

Pokorny, A. D. Characteristics of forty-four patients who subsequently committed suicide. *AMA Arch. gen. Psychiat.*, 1960, **2**, 314–323.

Robins, E., Gasner, S., Kayes, J., Wilkinson, R. H., & Murphy, G. E. The communication of suicidal intent: A study of 134 consecutive cases of successful (completed) suicide. *Amer. J. Psychiat.*, 1959, **115**, 724–733.

Rosen, A. Detection of suicidal patients: An example of some limitations in the prediction of infrequent events. *J. consult. Psychol.*, 1954, **18**, 397–403.

Sakheim, G. A. Suicidal responses on the Rorschach test: A validation study. *Dissertation Abstr.*, 1954, **14**, 1253–1254.

Shneidman, E. S., & Farberow, N. L. *Clues to suicide*. New York: McGraw-Hill, 1957.

Wilson, K. V. A distribution-free test of analysis of variance hypotheses. *Psychol. Bull.*, 1956, **53**, 96–101.

(Received June 14, 1960)

Journal of Consulting Psychology
1961, Vol. 25, No. 4, 316-323

CONCURRENT AND CONSTRUCT VALIDITY OF DIRECT AND INDIRECT MEASURES OF DEPENDENCY

MARVIN ZUCKERMAN [1]
Institute of Psychiatric Research, Indianapolis, Indiana
EUGENE E. LEVITT, AND BERNARD LUBIN [2]
Indiana University Medical Center

Research on acquiescence response set (Zuckerman, Norton, & Sprague, 1958) and suggestibility (Zuckerman & Grosz, 1958) led to an interest in the possible relatedness of these variables to the trait called "dependency." In assessing this trait, peer nominations, objective and projective tests had been used. However, anyone working in the field of personality plunges into chaos when he faces the problem of measurement. One finds that widely differing techniques claim to measure the same hypothetical variable, but in fact do not correlate with each other. The answer to this lack of construct validity given by some assessment theorists (Leary, 1957) is that the tests are measuring "different levels of personality." The concept of levels is related to the idea of a continuum of consciousness-unconsciousness. The more direct tests (those which ask the testee to describe his own feelings and reactions) are presumably at the upper levels, while the less direct tests (those which are disguised as creative or perceptual tasks) tap the lower, "unconscious," levels. This distinction does not resolve the psychometric dilemma, for one must still demonstrate that a test at any level is measuring something more general than itself, even if this is indicated only by a correlation with other measures at adjacent levels. Furthermore, at some point, a test should relate to behavior as assessed by something other than another test. This may be observations by psychologists in life situations, measurements in controlled miniature situations, or descriptions of the subject made by persons who have had a chance to observe him over some period of time.

The study reported here represents an attempt to bring some order to measurement of one construct: dependency. An attempt was made to: conceptualize the dimensions of the construct; select a battery of tests covering the range of the direct-indirect continuum and representing some of the current methodologies of test development; score these tests within the same conceptual framework; assess their validity in two ways: (*a*) concurrent validity, by comparing all tests against an external criterion, a rating by peers; (*b*) construct validity, by factor analyzing the correlations between all tests and peer ratings. It was assumed that the largest factor to emerge from a factor analysis of all variables would represent a broad dependency factor, and that loadings of the tests on this factor could be taken as an indication of their construct validity. Of course, secondary factors also might have some bearing on construct validity. In fact, it was hoped that the factor analysis might help conceptualize the dimensions of the broad construct, dependency, beyond the initial working hypotheses.

The peer rating was selected as the single criterion to use for concurrent validity because it was closest to behavioral description in a general sense. It should be noted that the

[1] Now at Department of Psychology, Brooklyn College, Brooklyn 10, New York.

[2] The authors are indebted to Beatrice Barrett, Robert A. Wagoner, Harry Brittain, and George Petoe for their assistance in judging and scoring test records.

measurement of personality from this source depends on the effect that the subject's behavior has on her peers, and does not necessarily have a relevance to the subject's own image of herself or her underlying motives and defenses.

Method

The concept of dependency used in this study stems from Horney's (1945) description of the "compliant" or "moving-toward-people" personality. Since some of the tests we considered using were scorable within the Murray (1938) need system, it was convenient to translate Horney into Murray. Horney delineated three traits of the compliant personality: a marked need for affection and approval from others (Succorance), a tendency to subordinate himself to others and to inhibit assertiveness and criticality (Deference and Abasement), a tendency toward self-blame and guilt (Abasement). Murray needs at the opposite poles are: Autonomy and Dominance.

Subjects

The initial subjects were 78 sophomore student nurses living together in a dormitory in the Indiana University Medical Center. They were acquainted with one another for about 5 months previous to the time when tests were given. Six of this initial group were later excluded from the study because their tests were incomplete or because their Consistency score on the EPPS was less than 9. The remaining group of 72 subjects were almost all between 19 and 20 years of age. Their average ACE score was 58.6 (national norms percentile), indicating slightly higher than average intelligence, relative to college students. The group's mean scores on the Edwards Personal Preference Schedule variables were compared with Edwards' (1954) means for 749 female college students. Critical ratios higher than 2.00 were obtained for differences on 6 of the 15 scales. The student nurse group scored significantly higher on Abasement, Intraception, Nurturance, and Endurance, and significantly lower on Dominance and Change. The differences on Dominance, Abasement, and Nurturance scores replicate differences found between a previous student nurse group from this school and Edwards' college females (Zuckerman, 1958). Additional comparisons were done on the Navran Dependency scale (1954) and Gough Dominance scale (Gough, McClosky, & Meehl, 1951). The student nurse group scored significantly higher ($p < .001$) on the Navran Dependency scale than a group of 200 normals described by Navran (1954). The average score of the nurse group on the Gough Dominance scale fell between the average scores reported (Gough et al., 1951) for high and low Dominance groups, but fell much closer to the mean for the low Dominance groups. In sum, it would appear that this group of subjects differs from the usual female college student group in various personality traits. The possible bearing of these differences on our results will be discussed at the end of this paper.

Peer Ratings (PR)

The items from four dimensions of Leary's Interpersonal Check List (Leary, 1957) were adapted to form three eight-point bipolar scales providing description at each point and fitting the hypothesized dimensions of the construct. These scales are given below:

I. Pride-Shame

1. Expects everyone to admire her
2. Always giving advice, acts important, tries to be too successful
3. Makes a good impression, often admired, respected by others
4. Well thought of
5. Able to criticize herself
6. Apologetic, easily embarrassed, lacks self-confidence
7. Self-punishing, shy, timid
8. Always ashamed of herself

II. Dominant-Submissive

1. Dictatorial
2. Bossy, dominating, manages others
3. Forceful, good leader; likes responsibility
4. Able to give orders
5. Can be obedient
6. Usually gives in, easily led, modest
7. Passive and unaggressive, meek, obeys too willingly
8. Spineless

III. Independent-Dependent

1. Egotistical, conceited, cold, unfeeling
2. Boastful, proud, self-satisfied, snobbish, thinks only of herself, shrewd, calculating, selfish
3. Independent, self-confident, self-reliant, can be indifferent to others, business-like
4. Self-respecting, able to take care of herself
5. Grateful, appreciative
6. Often helped by others, admires and imitates others, very respectful, very anxious to be approved of, accepts advice readily, trusting, and eager to please
7. Dependent, likes to be taken care of, easily fooled, wants to be led
8. Clinging vine, will believe anyone

The subjects were asked to rate every nurse in the group whom they felt they knew or had observed enough to rate. They rated a peer by circling one of the numbers from 1 to 8 on each of the three scales. A combination score was obtained by summing each subject's ratings for each peer that she rated.

Tests

Six basic kinds of instruments were used. In order of what was felt to be their directness, we have:

1. Self-Ratings (SR): All subjects made self-ratings on the same scales that they used to make peer ratings.

2. True-False Questionnaires (Q):

a. Gough (1951) Dominance scale, developed from the MMPI using an empirical method, or item selection based on criterion groups

b. Navran (1954) Dependency scale, developed from MMPI using content validity, or item selection by clinical judges

In addition to the scores derived from the separate questionnaires, a combination score suggested by Ullman (1958) was formed by counting items on both tests scored for submissiveness and dependency and eliminating overlapping items. Ullman called this combined scale "Lack of Self-Assertion."

3. Forced-Choice Questionnaire: Edwards Personal Preference Schedule (EPPS) (Edwards, 1954). It is assumed that the elimination of the social desirability factor makes this kind of forced-choice questionnaire less direct than the usual true-false questionnaire, since the subject is less able to choose his responses in conformity to a more obvious stereotype of "adjustment." Furthermore, the pairing of 15 needs probably makes it more difficult for the subject to grasp what is being measured than when all items pertaining to a particular need are presented in one scale. The only scales actually used were those relevant to our construct: Deference, Succorance, Abasement, Autonomy, and Dominance. A combination, or ratio, score was formed by converting the raw scores to Edwards' standard scores and taking the ratio of Deference plus Succorance plus Abasement to the total sum of all five scores.

4. Sentence Completion Test: Rohde's (1957) Sentence Completion Test (SC) was used since the manual describes scoring in the Murray need system. The SC often is classed as a projective test because it is free-response. But it is more direct than other projective tests because it asks the subject to describe his own feelings. The items were scored for the same variables as were scored in the EPPS, and the same ratio score was used as a combination score. The three experimenters scored the test one item at a time over all subjects. An initial discussion of the range of responses on the particular item was followed by independent scoring and then acceptance of scores with a minimum consensus (two out of three agreement on the basic score).

5. TAT: Ten cards from the TAT (Cards 2, 3GF, 4, 6, 7BM, 7GF, 9GF, 10, 12M, 18GF) were presented to the group using an opaque projector. Subjects were allowed 5 minutes per card to write their stories. The three experimenters scored the stories one story at a time over all subjects. Each experimenter gave the story an initial score, and disagreements were resolved by conference technique. The five needs scored were the same as in the EPPS and SC, and a ratio score was formed in the same manner as on these two previous tests.

6. Rorschach: The Rorschach was group administered using an opaque projector to project the cards on a screen. Subjects were allowed 3 minutes to write down the different things they could see on a card. A scoring system for Dependency content was adapted from that of DeVos (1952). After some practice scoring, it was found necessary to eliminate some of his more ambiguous categories. Two assistants independently scored the content of the 72 records. The correlation between scorings was .83. The frequency scores of the two scorers were averaged for each subject, and this average was divided by the subject's total number of responses to get a Dependency percentage score.

Results and Discussion

Peer Rating Reliabilities and Intercorrelations

The number of peers rating each subject ranged from 37 to 74 with a median of 57. Reliabilities of the peer ratings were calculated using a formula devised by Horst (1949). The resultant reliability coefficients for PR_I, PR_{II}, PR_{III}, and the Sum PR, were .94, .96, .94, and .96. The correlations between the three PRs were close to the maximum limit set by their reliability: .92, .91, and .92. Apparently, the subjects made no distinctions between the three scales in rating their peers. However, the halo effect cannot be attributed to a general positive-negative reaction because the "negative," or "undesirable," descriptions are at both ends of the scales. A general evaluation might affect degree, but could not affect direction of the ratings, i.e., the rater still had to decide whether she did not like the ratee because the ratee was too dependent or too independent. However, this "across-scales" halo effect precluded using the PR to reveal possible dimensions of the general construct. Therefore, only the Sum PR was used in all comparisons with other measures. This measure was averaged across judges for each subject. The distribution of these average PRs was essentially normal.

Self-Ratings: Intercorrelations

The subjects showed more variation between the three scales in rating themselves, for the correlations between SRs on the three scales were .44, .37, and .47. It was therefore possible to analyze these scales in a combination score *and* singly.

Analysis of Combination Scores

Table 1 lists the intercorrelations of the sum and ratio scores for each of the basic

techniques. The correlations between each of the tests and the PR can be seen in the first row across. These correlations tend to fall off as a function of the postulated indirectness of the test. The self-ratings yielded the highest correlation, the questionnaires of both types and the SC test were intermediate, and the TAT and Rorschach were lowest. Only the three more direct tests correlated significantly with the criterion, the SC test correlated just below the .05 level, and the TAT and Rorschach correlated close to zero with the criterion. The combined indices did not mask any high relationships between individual scores and the criterion, although in some cases one could have done just as well or slightly better using a single variable. For instance, the Gough Dominance scale correlated more highly (−.33) when used alone than when used in combination with the Navran Dependency scale. While the SC ratio did not correlate significantly with the PR, the SC Autonomy score did correlate significantly (−.33).

It can also be seen in Table 1 that all of the three most direct techniques were significantly intercorrelated. The SC test correlated significantly with the EPPS, and approached significance in its correlations with the other more direct techniques. The TAT and Rorschach did not correlate significantly with each other or any of the other techniques. Another result seen in this table is the fact that the highest positive correlations are on the diagonal of the matrix and that correlations tend to drop off or become negative in either direction from the diagonal. This means that each test tends to vary most directly with the tests closest to it on the postulated continuum of "directness-indirectness." This finding offers support for the original placement of these tests in this continuum. The SC test seems to have more in common with the direct objective tests than with the indirect free-response tests.

Another analysis of TAT and Rorschach was undertaken in order to check the possibility that a more global approach to the indirect techniques might yield greater concurrent validity. The TAT and Rorschach protocols of the 10 highest and 10 lowest subjects on the Sum PR criterion were selected for this purpose. These protocols were given to two experienced clinical psychologists along with the rating instructions used by the peers in making their ratings. They were asked to predict which subjects were rated as dependent and which as independent by their peers. They were informed that one-half of the group of 20 fell into each of these categories, and were asked to make their sortings using the same distribution. Quantitative scores involved in the comparisons were dichotomized at the medium and comparisons were made using exact probability tests for 2×2 tables.

TABLE 1

INTERCORRELATIONS OF SUM AND RATIO SCORES

Technique	SR	Q	PPS	SC	TAT	Ror	ACE
PR	.44**	.28*	.24*	.22	.05	.02	−.31**
SR		.37**	.37**	.23	−.22	−.12	−.14
Q			.52**	.18	−.10	−.19	−.25*
EPPS				.39**	.09	−.13	−.30*
SC					.11	−.17	−.27*
TAT						.08	−.10
Ror							.17

* Significant at or below .05 level.
** Significant at or below .01 level.

Using the TAT, the judges did not demonstrate significant agreement between themselves (60%), or with the PR criterion (40% and 70%), or the TAT ratio (50% and 60%). The lack of agreement with the latter score may be either a function of the unreliability of the judgments or a basic lack of comparability between the global and more atomistic methods involved.

Using the Rorschach, the agreement between judges was somewhat better (70%) but still short of significance. Agreement between both judges and PR was 50% or exactly chance probability. Agreement between both judges and the dichotomous classification based on the Rorschach Dependency percentage score was 80%. The probability of either of these judgments occurring by chance was less than .02. It would appear from this analysis that global judgments based on the Rorschach were related to the scores derived from the more atomistic scoring method, but were no more valid than those scores.

Factor Analysis of the Combination Scores

The matrix contained in Table 1 was factor analyzed using a principal components

program on the IBM 650. One large factor accounted for 92% of the variance contained in the communality estimates derived from multiple correlations on each variable.[3] The significant loadings (significance arbitrarily set at .30) on this General Dependency factor were as follows: PRs (.52), SRs (.59), Q (.61), EPPS (.68), and SC (.46). TAT and Rorschach had loadings of negligible magnitude (−.02 and −.23). The ACE had a significant negative loading (−.45).

The conclusions are essentially those drawn from the matrix. The three more direct tests demonstrate the greater validity, the SC test is intermediate, and the two most indirect tests demonstrate no validity. The loading of intelligence (ACE) could be interpreted in one of two ways: (*a*) intelligence is an extrinsic factor influencing the measures through response sets which are also extrinsic to the trait dependency, (*b*) intelligence is a factor which is intrinsically related to dependency and its evaluation in others. Although Hypothesis *a* would explain the correlation between ACE and the peer ratings (see Table 1), it is hard put to explain the correlation between ACE and EPPS, a test carefully controlled for the usual response sets, or the correlation with SC, a free-response test. It must be remembered that a high intelligence level gives one a greater potentiality for independence particularly in the academic and work situations.

Factor Analysis of the Individual Scores

All of the single variable scores were intercorrelated, and the resulting 23 × 23 matrix was factor analyzed using Indiana University's IBM 650. Five factors were found to account for 97% of the variance contained in the estimated communalities.[4] These factors were rotated to an orthogonal solution using the normalized varimax IBM program.[5]

Factor I was the most general Independence-Dependence factor obtained. Loadings of over .30 and in the expected direction were found for the Sum PR, all of the SRs, the Gough Dominance scale, all of the EPPS scales (with the exception of Abasement), and Autonomy in the SC test. Loadings in the unexpected direction were obtained from Abasement in the SC and TAT tests. These two loadings are perhaps explainable by the fact that the total score of all five variables was not partialed out of the scores for single variables. These total scores (measures of the intensity or scorability of the free-responses) were found to correlate negatively with many of the dependency measures, and high and positively with the Abasement variable in both tests.

As in the previous factor analysis, the more direct measures seemed to have the highest loadings on the most general dependency factor.

Factor II might be labeled Dominance vs. Abasement. It had positive loadings from the Gough Dominance and EPPS Dominance scales, and negative loadings from the Navran scale, EPPS Abasement, and SC Abasement scales. An examination of the Navran scale reveals that most of the items do have an abasement type of content with a lesser number of succorance items. EPPS Succorance also had a negative loading on this factor. The factor seems to involve assertion, leadership, and self-confidence at one pole; and yielding, resignation, inferiority feelings, and feelings of depression and fear at the other. The factor might relate to Cattell's (1957) Dominance vs. Submission factor, but it also seems to involve an emotionality element.

Factor III might be labeled Autonomy vs. Deference. It had positive loadings from the

[3] A copy of this factor matrix has been deposited with the American Documentation Institute. Order Document No. 6756 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

[4] Copies of the correlation matrix and the rotated factor matrix have been deposited with the American Documentation Institute. Order Document No. 6756 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

[5] Users of IBM rotation programs may be interested to know that the quartimax program was also used in order to compare it with the varimax program. Both programs yielded substantially similar results, and differences in loadings were minimal.

SC Deference and TAT Deference scores, and negative loadings from the SC Autonomy and TAT Autonomy scores. In addition, the TAT Succorance score had a negative loading and the ACE loaded positively on the factor. This factor is interesting in that it is derived entirely from two projective, or free-response, techniques, and has some obvious construct consistency.

Factor IV might be labeled Succorance. It had positive loadings from the Navran scale, EPPS Succorance, and SC Succorance. A negative loading was obtained from EPPS Autonomy. A puzzling positive loading was obtained from TAT Dominance.

Factor V was practically uninterpretable. It had negative loadings from SC Dominance and Autonomy on the one hand and TAT Abasement and Rorschach Dependency % on the other. In a subsequent analysis, this factor was collapsed and its variance distributed among the first four factors. This operation did not change the composition of these first four factors.

Beyond Factor I, General Dependency, the factors did not show extensive breadth across tests. Construct validity seems to span three tests at the most. Usually, tests which are closer in their degree of directness share some common variance on comparable variables.

The last factor analysis also exposed some of the obvious weaknesses of the tests. Although all three of the self-ratings loaded on the general dependency factor they did not discriminate among the other dimensions of the construct. Although the Navran scale is called "Dependency" it seems to be more of an abasement scale. The Rorschach had only one loading of any magnitude, and this was on a weak and uninterpretable factor. Proponents of the Rorschach might take some heart from an interesting side-finding in an unpublished study by Levitt and Zuckerman comparing the characteristics of Volunteers and Nonvolunteers for a hypnosis experiment, and using this sample of subjects. The Rorschach Dependency score was the only one which distinguished these two groups with the Volunteers scoring higher on it ($p < .01$).

But in general, whether we consider concurrent validity (correlation with the peer rating) or construct validity (loading on a general dependency factor), the more direct tests seem to do better than the less direct tests. There is a certain irony in these results if one considers the time and effort expended on them. The self-ratings required only a few minutes to administer, and a technician could record one subject's score in less than a minute. The TAT took about an hour to administer and about an hour each for three trained psychologists to score one subject's record. If one starts calculating cost in terms of return (validity) a kind of moral can be drawn: try out simple techniques before resorting to complex techniques. This is not the first study to suggest that empirically developed, objective, and simple psychometric instruments may do better, just as well, or just as poorly as more complex free response techniques.

There are, however, important reservations to forming sweeping conclusions from the results of this study:

1. The concurrent validity of self-ratings may be a function of their greater similarity in form to the peer ratings. However, one would be rather foolish not to make self-ratings as congruent as possible with the criterion being predicted.

2. The peer and self-rating methods are lacking in that they do not discriminate dimensions within the construct, and are vulnerable to halo effect.

3. The results may not have applicability for other kinds of constructs. Simple techniques such as the Taylor scale (1953) and an Affect Adjective Check List devised by the senior author (Zuckerman, 1960) seem to work well in measuring "anxiety," but one wonders what the results would be with a less socially acceptable symptom like "hostility." One would expect that the less socially acceptable constructs would be more difficult to measure with direct techniques, but it does not follow from this that indirect techniques would do better.

4. The results may be in some part a function of the particular type of subjects used. As a group, the Indiana student nurse seems to be more abasing and nurturant, and less dominant than Edwards' (1954) college females on the EPPS. Perhaps dependency (particularly the submissive component) is

more socially acceptable in terms of the "nurse" ego-ideal. This particular group of nurses is also brighter than average, and higher on the EPPS Intraception score. Direct techniques may be less useful with duller or less "self-analytic" groups.

Allport (1953) has suggested that direct measures of motivation will be more effective than indirect measures with normal subjects, while indirect, or projective, techniques may be useful in assessing the maladjusted. A conservative conclusion from this study would be a simple underlining of Allport's (1953) statement: "A psychodiagnostician never should employ projective methods in the study of motivation without at the same time employing direct methods" (p. 111).

Summary

The purpose of the study was to test the validity of a range of direct and indirect tests against a peer rating criterion (concurrent validity) and the factors derived from a factor analysis of all measures (construct validity). Another purpose was to clarify the dimensions of the broad construct, "dependency."

The subjects were 72 student nurses. They rated themselves (self-rating) and each other (peer rating) on three bipolar, eight-point scales derived from Leary's Interpersonal Adjective Check List. All subjects took the Gough Dominance and Navran Dependency questionnaires, the Edwards Personal Preference Schedule, the Rohde Sentence Completion Test, and a group administered TAT and Rorschach. The EPPS, SC, and TAT tests were scored for five relevant Murray needs: Autonomy, Dominance, Succorance, Abasement, and Deference. The Rorschach content was scored using a system adapted from DeVos. Combination scores were obtained by: summing the three peer ratings, summing the three self-ratings, combining the two true-false questionnaires, and using ratios combining the five Murray needs on the EPPS, SC, and TAT. Combination scores were correlated with peer ratings and with each other, and the resulting matrix was factor analyzed. Individual scores on all tests were analyzed in a second factor analysis.

Using combination scores, the self-ratings, questionnaires, and EPPS scores correlated significantly with the peer ratings, while the SC, TAT, and Rorschach did not. The magnitude of the validity correlations tended to drop as function of the indirectness of the tests. More global judgments of the TAT and Rorschach, using extreme groups on the peer rating distribution, did not indicate any greater concurrent validity for these tests. A factor analysis of the combination scores yielded one large factor called General Dependency. All of the four most direct tests showed moderate to high loadings on this factor while the less direct tests (TAT and Rorschach) showed negligible loadings. A factor analysis of the individual scores yielded four interpretable factors: (*a*) General Dependency, (*b*) Dominance vs. Abasement, (*c*) Autonomy vs. Deference, (*d*) Succorance. Factors *a*, *b*, and *d* were composed mainly of individual scores from the peer and self-ratings, questionnaires, EPPS, and SC tests. Factor *c* was composed entirely of scores from the SC and TAT tests.

In general, the more direct measures of dependency demonstrated the greater validity of both types, but the conclusions about the relative validity of the two types of tests are limited by the form of the measures, the particular type of subjects used, and the particular construct investigated.

References

Allport, G. W. The trend in motivational theory. *Amer. J. Orthopsychiat.*, 1953, **23**, 107–119.

Cattell, R. B. *Handbook for the Sixteen Personality Factor Questionnaire*. Chicago: Psychometric Affiliates, 1957.

DeVos, G. A quantitative approach to affective symbolism in Rorschach responses. *J. proj. Tech.*, 1952, **16**, 133–150.

Edwards, A. L. *Edwards Personal Preference Schedule: Manual*. New York: Psychological Corporation, 1954.

Gough, H. G., McClosky, H., & Meehl, P. E. A personality scale for dominance. *J. abnorm. soc. Psychol.*, 1951, **46**, 360–366.

Horney, K. *Our inner conflicts*. New York: Norton, 1945.

Horst, P. A generalized expression for the reliability of measures. *Psychometrika*, 1949, **14**, 21–23.

Leary, T. *Interpersonal diagnosis of personality*. New York: Ronald, 1957.

Murray, H. A. *Explorations in personality*. New York: Oxford Univer. Press, 1938.

NAVRAN, L. A rationality derived MMPI scale to measure dependence. *J. consult. Psychol.*, 1954, **18**, 192.

ROHDE, AMANDA R. *The sentence completion method.* New York: Ronald, 1957.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

ULLMAN, L. On the relationship between amount of hospitalization and self-assertion. *Amer. Psychologist*, 1958, **13**, 327. (Abstract)

ZUCKERMAN, M. The validity of the Edwards Personal Preference Schedule in the measurement of dependency-rebelliousness. *J. clin. Psychol.*, 1958, **14**, 379–382.

ZUCKERMAN, M. The development of an Affect Adjective Check List for the measurement of anxiety. *J. consult. Psychol.*, 1960, **24**, 457–462.

ZUCKERMAN, M., & GROSZ, H. J. Suggestibility and dependency. *J. consult. Psychol.*, 1958, **22**, 328.

ZUCKERMAN, M., NORTON, J., & SPRAGUE, D. Acquiescence and extremes sets and their role in tests of authoritarianism and parental attitudes. *Psychiat. res. Rep.*, 1958, **10**, 28–45.

(Received June 20, 1960)

Journal of Consulting Psychology
1961, Vol. 25, No. 4, 324–326

RESPONSE BIAS IN QUESTIONNAIRE REPORTS[1]

N. H. AZRIN, W. HOLZ, AND I. GOLDIAMOND[2]
Anna State Hospital, Illinois

If one defines psychology as the study of behavior, the direct measurement of behavior appears to be a minimal prerequisite to further analysis. Several alternatives to a direct measurement of behavior are commonly practiced. One such indirect method defines and measures the behavior in terms of the effect of that behavior upon the environment. For example, the measurement of reaction time typically is based upon the moment of closure of an electric switch or push-button. The simple fact of closure of the push-button does not guarantee that the movement of any one finger be involved since any finger or even the palm, wrist, arm, or leg might just as easily have been used. Such ambiguities in interpretation are easily overcome, and the experimenter can easily confirm his interpretation of the switch closure by occasionally or continuously observing the behavior directly.

A second alternative to direct behavioral observation is the interview or questionnaire procedure. Here the experimenter typically does not have any simple means of direct behavioral observation. Rather, the subject himself is expected to observe his own behavior and to describe it at some future date. The subject's reports usually cannot be evaluated by direct observation of the behavior being reported as was true of closing of the switch. The problem is often enhanced by the fact that the behavior being reported upon is basically unobservable by the experimenter by its very nature. This is true for the so-called "subjective reactions" as when an indiviual states that he feels hostile or afraid. In addition, it is quite likely that a report of one's own behavior will be modified considerably by the audience or experimenter to whom the report is being made. Other factors such as social acceptability may also be involved. The reply to the question "Did you cheat?" would probably be different if the interviewer were a classmate than if he were an instructor. For whatever the ultimate reasons may be, the individual being questioned may have a pre-existing tendency or bias in admitting to some statements and not to others. The present study was performed in order to study the influence of such response biases upon the reports of behavior obtained through a questionnaire.

A well-known study by Shaffer (1947) deals with the reports of combat flyers of their fears in combat. The reports had been obtained from these flyers by means of a questionnaire some 2 months following the termination of their combat experiences. This study has been widely interpreted as demonstrating that some behavioral reactions such as "soiling one's pants" are more indicative of fear in combat than are behavioral reactions such as "feeling nervous and tense." In order to evaluate the validity of this interpretation, 160 college freshmen and sophomores, including males and females, in five separate psychology and sociology classes were given a questionnaire. This questionnaire contained the same 15 "symptoms" reported in the original investigation by Shaffer. All of the questionnaires had the following directions:

> Imagine that you are a combat flyer who has flown many missions over enemy territory. Your commanding officer gives you the questionnaire below and tells you to fill it in. Fill in the answers keeping in mind what your commanding officer expects you to have felt.

Two forms of the questionnaire were used, however; half of each class of students re-

[1] This investigation was supported by a grant from the Psychiatric Training and Research Fund of the Illinois Department of Public Welfare.

[2] Now at Arizona State College, Tempe, Arizona.

ceived one form and half received the other form. One form stated after the first sentence:

> You have been extremely frightened on all of your missions and have experienced each of the symptoms below on every flight.

The other form stated:

> You have never been frightened on any of your missions and have not experienced each of the symptoms below on every flight.

One half of the students are thereby told that they have never experienced any of the listed symptoms, whereas the other half are told they have experienced all of the symptoms. Further, all students were instructed by the questionnaire to answer in terms of what is expected, regardless of what behavior is presumed to have occurred.

Results and Discussion

Table 1 presents the percentage of students who chose each symptom as occurring "often" or "sometimes." It will be recalled that the students had been told that all symptoms were experienced equally often. The only basis for checking some symptoms more than others was the specific instruction to keep in mind what answers are expected. Had there been no predisposition or response bias toward some symptoms, one should expect all symptoms to have been selected equally often. Certainly, no particular rank ordering of the symptoms should have emerged. The results of Table 1 demonstrate that the selection of symptoms does not follow a random distribution. Some symptoms were selected as much as six times as often as others.

Spearman rank-order correlations were performed to determine the consistency of this response bias among the students. It was found that the rank-order of symptoms for males correlated with that of females with $\rho = .88$. Similarly, the rank-order correlation of symptoms was .95 between those students who were told that they had experienced all the symptoms and those who were told that they had not experienced the symptoms. This high degree of similarity demonstrates that the response bias toward certain symptoms exists regardless of whether or not the symptom was alleged to occur.

TABLE 1

Reported Symptoms of Combat Fear of 160 College Students

During combat missions did you feel:	% of students stating "often" or "sometimes"
That your muscles were very tense	72
A pounding heart and rapid pulse	71
"Butterflies" in the stomach	67
Dryness of the throat or mouth	67
"Nervous perspiration" or "cold sweat"	61
Sense of unreality that this couldn't be happening to you	49
Easily irritated, angry, or "sore"	43
Need to urinate very frequently	42
Trembling	39
Unable to concentrate	36
Sick to the stomach	34
Right after a mission, unable to remember details of what happened	32
Confused or rattled	28
Weak or faint	25
That you have wet or soiled your pants	11

In order to determine whether the same response bias might have affected the reports of the combat flyers, a Spearman rank-order correlation was performed between the symptoms reported by the flyers and those reported by the students. It was found that the rank orders of the responses were highly similar ($\rho = .89$) between the students and the flyers. Nor was this relationship reduced for those students who were told that they had not experienced the symptoms ($\rho = .94$) as compared with those who were told that they had experienced all of the symptoms ($\rho = .90$). The statistical stability of this response bias is evidenced by the degree to which the rank-order in each classroom of students was correlated with the rank-order reported by the combat flyers: $\rho = .70, .82, .85, .90$, and $.92$.

The response pattern obtained from the students by means of the questionnaire is al-

most completely predictable on the basis of response bias. Therefore, it is quite likely that the same type of response bias operated on the combat flyers. Any conclusions concerning the actual symptoms must await study by a method that provides for a more direct and objective measurement.

The present findings may well be considered for their implications for the use of interview and questionnaire methods in general. Unless an objective and direct means of measurement is available, the questionnaire responses may be independent of the behavior being studied. A definitive method of determining the validity of the reports is the direct and objective measurement of the behavior being reported. Once such a direct measure is available, however, the very need for questionnaire reports is eliminated.

REFERENCE

SHAFFER, L. F. Fear and courage in aerial combat. *J. consult. Psychol.*, 1947, **11**, 137–143.

(Received June 21, 1960)

Journal of Consulting Psychology
1961, Vol. 25, No. 4, 327–336

THE MMPI AS A MEASURE OF CHARACTER STRUCTURE AS REVEALED BY FACTOR ANALYSIS[1]

JOSEPH C. FINNEY
Hawaii Department of Health, Honolulu

The psychologist who needs conceptual tools to understand and to help people, faces a dilemma. On the one hand he can accept and use dynamic concepts, usually psychoanalytic ones, including urges and defense mechanisms, which seem fairly satisfactory in working with patients. If he does so, it is with the disadvantage that he assumes some cause and effect relationships on slender evidence, and that he treats as quantities many things that he has no way of measuring. On the other hand, if he is hard-boiled, skeptical, operational, and parsimonious, he may confine himself to relationships experimentally or statistically established, and to quantities that can be accurately measured. If so, he hampers himself more seriously: his concepts are at a level too crude and simple to be of value in grasping and dealing with his patients' troubles.

There is, further, a variety of opinion about what quantities an objective psychological test should measure (Cattell, 1957a; Gough, 1957; Hathaway, 1951). Even factor analytic studies (Cattell, 1957b; Cook & Wherry, 1950; Cottle, 1950; Lingoes, 1960; Welsh, 1956; Wheeler, Little, & Lehner, 1951; Williams & Lawrence, 1954) have not agreed on what fundamental dimensions are uncovered. It seems desirable that tests aimed at variables of importance in clinical, including psychoanalytic work, be included in factor analytic studies. The recent development of programs by which factor analysis, including rotation to oblique simple structure, can be performed entirely by electronic computer, (Dickman, 1959; Pinzka & Saunders, 1954) makes it possible for the psychologist with minimal mathematical training and little time for computation to do such studies.

Aims of this Study

The battery of measurements used in this study evolved gradually from the clinical demands of a county mental health clinic. For nearly 4 years the author and other psychologists made blind interpretations from the MMPI scores of applicants for clinic services. These personality descriptions were compared with the reports of interviews, and attempts made to account for discrepancies. One result was to increase the psychologists' skill in drawing inferences from MMPI scores. Even so, there remained many areas, regarded as important in clinical interviews, which no standard MMPI scale seemed to measure. One by one, additional scales were added to fill the gaps.

The strongly felt need for a measure of strength of ego-ideal or conscience was met by Gough's (1957) responsibility scale. The clinical desire for measures of hysterical character (defense of repression) and paranoid character (defense of projection) was tentatively met by adding to the battery Wiener's (1948) "subtle" subscales of *Hy* and *Pa*, which sounded promising. The "ego functions" were

[1] The following materials have been deposited with the American Documentation Institute: Tables A through M, significant reference vector loadings, for factors comparable among groups; lists of items in new scales; intercorrelations of the 59 variables; successive latent roots of factors; complete tables of loadings of the first two centroid factors; *D* and inverse of *D;* correlations among oblique factors; complete tables of reference vector and primary factor loadings of oblique factors. Order Document No. 6757 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $2.75 for microfilm or $7.50 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

tapped by some of Gough's (1957) scales, such as *Ie* (intellectual efficiency) and *Do* (dominance). Finally, some still untouched areas, such as stinginess and warmth, were included by adding several unpublished experimental scales devised for the purpose by the author. With the aid of the supplementary scales, the psychologist interpreting blindly from test scores became able to deal with the aspects of personality that interested the clinical interviewers, and with a satisfactory degree of agreement.

The question arises, which of the many measures represent important basic aspects of personality? Cattell (1952) has contended that factor analysis with rotation to oblique simple structure eliminates the superficialities and reveals the fundamental underlying dimensions. It seemed appropriate to apply that method to the group of personality measures that were used.

Method

A questionnaire of 600 items was given routinely to all willing literate, nonpsychotic adults applying for services (whether as patients or as parents of child patients) at a community mental health clinic in Illinois, until 50 men's and 50 women's records were obtained (a process which took about 6 months). The questionnaire consisted of the booklet form of the MMPI, with some unscored and duplicating items omitted and some additional items included.

Tests were scored for 56 scales and two individual items. Intercorrelations were computed for 59 variables (the 58 test variables plus sex), using raw scores. The variables used were:

1. Sex (male or female)
2. Sex experience (A single item: "I've had sex relations with at least three different people in my life.")
3. Anti-Catholic prejudice (A single item: "I have no religious prejudice, but I think it would be bad for the country if a Catholic were elected President." The study was done in 1959.)
4. *L*
5. *F*
6. *K*
7. *Hs* (without *K* correction)
8. *HsK* (*Hs* with *K* correction)
9. *D*
10. *Hy*
11. *Pd* (without *K* correction)
12. *PdK*
13. *Mf*
14. *F* minus *K* (Gough, 1950)
15. *PaO* (Wiener's, 1948, "obvious" subscale of *Pa*)
16. *Pa*
17. *Pt* (without *K* correction)
18. *PtK*
19. *Sc* (without *K* correction)
20. *ScK*
21. *Ma* (without *K* correction)
22. *MaK*
23. *Si* (Drake, 1946)
24. *A* (Welsh, 1956)
25. *MA* (manifest anxiety, Taylor, 1953)
26. *Ie* (intellectual efficiency, Gough, 1957, MMPI items only)
27. *St* (capacity for social status, Gough, 1957, MMPI items only)
28. *Es* (ego strength, Barron, 1953)
29. *Do* (dominance, Gough, 1957, MMPI items only)
30. *Re* (responsibility, Gough, 1957, MMPI items only)
31. H_1 (an experimental scale on hostility, 117 items)
32. H_2 (an experimental scale on hostility, 146 items)
33. *Sc* minus *Ma*
34. W_2 (an experimental scale on warmth, 81 items)
35. *ODy* (an experimental scale on optimistic dependency, 9 items)
36. *Dy* (dependency, Navran, 1952)
37. *Ph* (an experimental scale for phobia, 12 items)
38. *Sd* (an experimental scale for sadism, 12 items)
39. *Ob* (an experimental scale for obsession, 16 items)
40. *An* (an experimental scale for anal or compulsive character, containing Scales 41, 43, 44, and 45 and a few additional items, 48 items)
41. *Rig* (rigidity-flexibility, Gough, 1957, reduced to 15 items)
42. *AuF* (the Berkeley F Scale, Adorno, Frenkel-Brunswick, Levinson, & Sanford, 1950, reduced to 15 items)
43. *Or* (an experimental scale for orderliness, 9 items such as: "I like to have my clothes clean at all times.")
44. *Sti* (an experimental scale for stinginess, 6 items such as: "I have to admit I hate to see money wasted.")
45. *Stb* (an experimental scale for stubbornness, 10 items such as: "People can push me just so far, and then I have to take a stand.")
46. *Reb* (an experimental scale for rebelliousness, 20 items)
47. *Sub* (an experimental scale for submissiveness, 26 items)
48. *Ap* (accepted passivity, Harris, unpublished)
49. *Id* (an experimental scale intended to measure inner direction versus other direction—concept of Riesman, Glazer, & Denney—21 items)
50. *De* (delinquency, Gough & Peterson, 1952, renamed socialization, Gough, 1957, MMPI items only)
51. *Fm* (feminine masochism, Hecht, 1952)
52. *PaS* (Wiener's, 1948, subtle subscale of *Pa*)
53. *R* (Welsh, 1956)

54. *HyS* (Wiener's, 1948, subtle subscale of *Hy*. Our selection followed a somatic-nonsomatic division, and differed slightly from Wiener's in including Items 137 and 179 and excluding Item 190. Little's *Dn* scale (Little & Fisher, 1958) is also virtually identical with *HyS*)

55. *Rep* (an experimental scale for repression, containing the 29 *HyS* items plus 17 others)

56. *Em* (an experimental scale for embarrassment, 13 items)

57. *Fe* (femininity, Gough, 1957, MMPI items only)

58. *Sc* minus *Pt*

59. *Hy* minus *PtK*

A Thurstone centroid factor analysis was performed by electronic computer. Rotation to oblique simple structure was also performed entirely by electronic digital computer, using a standard program (Dickman, 1959; Pinzka & Saunders, 1954) so that no human judgment entered into the selection of the angles of rotation. Both reference vector loadings and primary factor loadings were obtained.

Intercorrelations and factor analyses were performed separately on the group of 50 men (hereinafter called Group M), the group of 50 women (called Group W), and the combined group (Group C).

A question may be raised about the use of scales with common items. The standard MMPI scales have many items common to three or more scales, and the subscales used here compound this practice. The entire *PaS* scale of Wiener, for instance, is included in the *Pa* scale. Welsh (1956) has considered this problem serious enough to eliminate all common items before calculation. Any correlation, however, may be considered as the result of hypothetical common elements (McNemar, 1949, pp. 117–118), and it has been pointed out (Wheeler, 1951) that the problem of correlations among scales is the same whether it is due to physically identical items or to common meanings in different items.

Results and Discussion

Centroid Factors

Using the criterion of continuing to extract factors as long as the latent root exceeds 1.0, 13 factors were found in Group M, 12 in Group W, and 12 in Group C. The 12 to 13 factors extracted accounted for 86.4% of the variance in Group M, 85.3% in Group W, and 82.0% in Group C.

The Thurstone centroid factor analysis produced, in all three groups, a first factor most heavily loaded on uncorrected *Pt* (reference vector loadings .90, .95, .94, dropping to .73, .86, and .85 with *K* correction), Welsh's *A* (.87, .92, .91), Navran's *Dy* (.89, .92, .92), uncorrected *Sc* (.86, .93, .90), *F*-minus-*K* (.75, .89, .81), and Taylor's *MA* (.85, .89, .89); and with substantial negative loadings on Gough's *Ie* (−.76, −.85, −.85) and Barron's *Es* (−.78, −.85, −.85) (Table A).[2] This factor accounted for 28.7% of the variance in Group M, 36.6% of the variance in Group W, and 33.3% of the variance in Group C.

In all three groups, the second factor was most heavily loaded in the positive direction with Welsh's *R* (.45, .60, .68), Harris' *Ap* (.40, .47, .53), Gough's *Re* (.53, .32, .62), Drake's *Si* (.53, .33, .38), and experimental scales for embarrassment and submissiveness; and loaded negatively with *Ma* (−.66, −.51, −.61) and with experimental scales for stubbornness and rebellion (Table B). The first two factors together accounted for 40.5% of the variance in Group M, 45.6% of the variance in Group W, and 44.2% of the variance in Group C.

The first two factors obtained by the Thurstone centroid method correspond closely with those previously obtained by Welsh (1956) using a similar procedure. Welsh's scales *A* and *R* (1956) were designed to measure these dimensions.

Thurstone centroid factor analysis defines dimensions with the maximum common variance, i.e., in which the greatest number of measures will vary together, or in agreement with each other. That procedure seems designed to maximize the influence of "response set." Examination of the evidence suggests that this is indeed the case, and that Welsh's (and the present) first two centroid factors, *A* and *R*, are very largely composed of response set.

One much studied response set is the tendency to present oneself as, on the one hand, psychologically healthy and normal, or, on the other hand, as emotionally upset and in need of help. This aspect has been described by the Minnesota group as "faking good" or "faking bad," by Gough (1950) as "dissimulation," by Edwards (1957) as "social desirability," and by DeSoto and Kuethe (1959) as "the set to claim undesirable symptoms." This falsifying tendency, whether conscious or not, is

[2] Tables A through M are included in the materials available through the American Documentation Institute.

believed to influence questionnaire test scores seriously, the more so in tests with much face validity or obviousness. Repeated efforts have been made to eliminate this source of variance from clinical tests, either at the source (Edwards) or by suppressor variables (the earliest of such efforts being Meehl's *K* correction—Meehl & Hathaway, 1946).

Evidence suggests that the first centroid factor consists largely of this response tendency, the willingness or unwillingness to admit psychological sickness. The evidence consists of the high loadings on this factor of several scales already known to measure this tendency. One such scale is Gough's (1950) *F*-minus-*K* dissimulation index, with reference vector loadings of .75, .89, and .81 on this factor. Another is uncorrected *Pt*, much of whose variance DeSoto and Kuethe (1959) have shown to represent a "set to claim undesirable symptoms," and which has loadings of .90, .95, and .94 on this factor.

Taylor's and Welsh's "anxiety" scales are very highly correlated with uncorrected *Pt* (e.g., in Group C, .92 and .93, respectively), and may also be described as consisting largely of this response set. Both Welsh's and Taylor's scales consist of items recognizably "sick" in their frank verbal content—items that might easily be avoided by a person seeking to appear "normal." With these considerations, it is easy to understand the failure of Taylor's scale (Kendall, 1954; Lauterbach, 1958) to be validated as a measure of anxiety.

It is noted that *K* correction, which increases the validity of five clinical scales (Meehl & Hathaway, 1946), invariably decreased their loadings on this factor. If a factor measures genuine pathology, the suppressor variable *K* should increase factor loadings.

The second centroid factor showed no consistent relationship to measures of dissimulation. This is consistent with Wiggins and Rumrill's (1959) finding of lack of relationship between dissimulation and Welsh's *R* scale.

A second response set that has been widely studied is the tendency to answer "true" or "false." (A good review of the literature on response sets has been done by Wiggins—1959; Wiggins & Rumrill, 1959—and will not be repeated here.) There are some evidences to suggest that the second centroid factor consists largely of this "response bias."

Welsh derived his *R* scale (1956) to measure his second centroid factor. The present study replicates Welsh's in finding substantial loadings (.45, .60, .68) of the second centroid factor on the *R* scale. All 40 items of the *R* scale are keyed false.

Couch and Keniston (1960) investigated the True or False response tendency in detail, and described the personality characteristics of "yeasayers" and "naysayers" in terms strikingly like those used by clinicians for low and high *R* scorers. Couch's "overall agreement score," however, gives evidence of representing not only yeasaying but also the other type of response set; its correlations with MMPI variables and with 16 PF resemble the second order factor that Cattell calls "anxiety" (perhaps better named "willingness to admit sickness") and Welsh's and the present first centroid factor. This is not surprising, since in most questionnaires the Yes answers tend to be the sick sounding ones. The present study may have succeeded in separating these tendencies only in an orthogonal factor analysis.

As a check on yeasaying and naysaying, the loadings of the scales on the second centroid factor, for Group C, were correlated with the percentage of true keyed items in the scales. Single-item variables as well as *Mf* (spurious for mixed-sex groups because of scoring differences) were eliminated. The result is only suggestive, since scales contained common items, but the correlation of $-.60$ is at least consistent with the view that the second centroid factor is related to a Naysaying response bias. (The first centroid correlated .37 with "percentage true," no doubt a result of the aforementioned tendency of True answers to be sick sounding.)

To identify the first two centroid factors (and Welsh's *A* and *R* scales) as largely response tendency is not to discard them as trivial. The clinician values a measure of the first factor, willingness or unwillingness to admit psychological sickness, as a sign of a patient's motivation for treatment. And Couch and Keniston (1960) have shown personality correlates (id and impulsiveness versus super-

ego and inhibition) for the second response set.

Obliquely Rotated Factors, Cross-Validated

General. Both of the response set factors disappeared on rotation to oblique simple structure, and a quite different set of factors appeared. Indeed, some of the scales most heavily loaded with the first centroid factor, such as Welsh's and Taylor's anxiety and *F*-minus-*K*, were remarkable for showing little or no loading on any of the oblique factors.

This result is to be expected. A response set, almost by definition, influences many scales. The centroid method, as pointed out above, is designed to maximize the influence of response sets. Rotation to oblique simple structure, however, is designed otherwise. The theory behind this procedure assumes that any one scale should be related to only a very few fundamental dimensions, and hence this method seeks factors each having near-naught loadings on as many scales as possible. A vector such as the first centroid, which has substantial loadings on most of the scales used (Table A) is avoided. It should be impossible for this procedure to define a factor substantially loaded with a widespread response set. Cattell (1952) has given reasons for thinking that rotation to oblique simple structure brushes the superficialities away and discovers the fundamental underlying dimensions.[3]

The rotation was performed exclusively by electronic digital computer in order to avoid the possibility that the investigator's preconceived notions of personality dimensions might influence the type of factors found.

Examination of the 12 to 13 factors extracted from each group showed that 5 or possibly 6 factors were common to all three samples. These are shown in Tables C, D, E, F, G, and H. The factors that were found in both Groups M and W (and also in the combined Group C) can be regarded as confirmed by cross-validation. The factors found in only one group may be regarded as having failed in cross-validation, though some of them may be genuine factors peculiar to one sex. Only a further cross-validation on a new sample can decide this question.

Factor 1: Anal compulsive character or reaction formation. One factor corresponded closely with Freud's 1908 description of the anal character (Freud, 1950). It contained substantial loadings on three experimental scales designed to measure Freud's "anal triad": orderliness (.61, .61, .61), stinginess (.34, .23, .41), and stubbornness (.40, .21, .30), respectively (Table C). These three scales contained no common items. Higher than any of these were the loadings on Gough's (1957) rigidity-flexibility scale (.70, .74, .70). A still higher loading (.73, .71, .78) was on an experimental scale designed to measure anal character, containing the four scales just mentioned.

An abbreviated form of the Berkeley authoritarian F Scale (AuF) also showed positive loadings on this factor (.32, .15, .24).

This factor seems clearly to correspond to the personality type known commonly as the anal or compulsive character. In addition to Freud's original paper in 1908, this type has been described by Jones (1913) and by Fenichel (1945). The concept of compulsive character has had widespread clinical acceptance and is recognized as an official diagnosis by the American Medical Association and the American Psychiatric Association (1952).

Despite clinical acceptance, there has hitherto been scant experimental evidence to support the concept of anal or compulsive character. In the one outstanding positive study, Sears (1942), dealing with ratings of fraternity brothers by each other, found (when halo effect was partialed out) positive correlations of .36–.39 among ratings of orderliness, stinginess, and stubbornness. His finding was all the more remarkable because the young men regarded orderliness as desirable and stinginess and stubbornness as undesirable.

The present finding of a strong relationship between the anal personality factor and Gough's rigidity scale is in accord with the

[3] It should be noted that this procedure does not eliminate response set from the items themselves. Hence an item analysis, using an oblique factor as criterion, may still produce a scale biased with response set, unless special precautions are taken. This is evidently the case with Cattell's 16 PF, in which 7 of the 16 scales are correlated both with *F* and also (in the reverse direction) with *K*, evidence of dissimulation effect (Karson & Pool, 1957).

common clinical impression of the compulsive person as "rigid" and suggests that the relationship between this factor and the experimental findings on perceptual rigidity (Luchins, 1942; Rokeach, 1948) should be explored. Since the experimental scales *Or*, *Sti*, and *Stb* have not been separately validated, Gough's rigidity scale must be regarded as the chief defining scale for this factor. It is noted that the rigidity scale has higher loadings than orderliness, stinginess, and stubbornness, and it is suggested that rigidity is the most central characteristic of the anal or compulsive factor.

Anal character was the factor that accounted for the largest portion of the total variance. The remaining factors are presented in arbitrary order. Wherever necessary, a factor from one of the patient groups has been reversed in sign, in all its loadings, to make it comparable with a factor derived from another patient group.

Factor 2: Hysterical character or repression. Another factor common to the three samples was one whose highest loading (among previously published scales) was on Wiener's subtle *Hy* scale or *HyS* (.58, .49, .58). It also had smaller positive loadings on *Hy* (.39, .16, .28) and *K* (.15, .13, .26), and negative loadings on *F* (−.18, −.33, −.22) (Table D). The highest loading (.69, .78, .70) was on *Rep*, an experimental scale designed to measure repression or hysterical character, and containing all of the *HyS* items with others.

This factor also seems to correspond to a character type long accepted in psychoanalysis (Fenichel, 1945). Rosenzweig (1945) described an "impunitive" reaction characteristic of hysterical persons, and a projective test for measuring it. McKinley and Hathaway (1944) found that persons with diagnosed conversion hysteria (physical symptoms such as headache and vomiting) answered certain questionnaire items in a naive Pollyanna-like manner. Wiener (1948) separated these items (*HyS*) from the full *Hy* scale. Eriksen, in a series of studies, related psychiatric diagnoses of hysteria, an MMPI measure (*Hy* minus *Pt* with *K* correction) and a perceptual measure of repression (Eriksen, 1952, 1954; Lazarus, Eriksen, & Fonda, 1951).

While hysterical conversion reaction is an official medical diagnosis (American Psychiatric Association, 1952), hysterical personality has never attained that status, probably because it is so common and is not considered sick enough.

This was the only oblique factor in which a dissimulation measure, *F*-minus-*K*, had a substantial loading (−.21, −.28, −.30), and it was in the paradoxical direction, such that faking good would appear to make one's score sicker. Other measures of dissimulation (*A*, Taylor *MA*, *Pt*) were unrelated to the factor. Consideration of the theory of repression suggests that it is genuine repression that is producing the appearance of dissimulation, rather than dissimulation contaminating the definition of the repression factor.

R. B. Cattell, who has kindly examined the present study, suggests (personal communication, 1960) that the hysterical character or repression factor found here may be the same as Cattell's Factor I, called "premsia." Factor II of Lingoes (1960) also resembles the present factor of repression or hysterical personality.

Each of the factors discussed so far, compulsive character and hysterical character, has a heavy loading on a previously published scale (Gough's rigidity and Wiener's *HyS*). In each case, there is an experimental scale (*An* and *Rep*, respectively) which includes the previously published scale along with other items. The *An* and *Rep* scales were attempts of the investigator to improve the measurement of the compulsive and hysterical personalities, and their higher factor loadings are evidence of a moderate degree of success in this endeavor.

One reason why *Rep* is a better measure of hysterical personality than *HyS* may be that it contains not only the *HyS* items of repression and denial, but also items designed to tap histrionic dramatization. Another reason may be its greater freedom from response sets. The *HyS* items are almost all keyed false, and, in the keyed direction, are "zero" (i.e., popular) responses that deny sickness and appear socially desirable (Wiener, 1948). The *Rep* scale is somewhat better balanced for both kinds of response set as well as for infrequency, and this may be why it is a purer measure of the hysterical character factor.

Factor 3: Paranoid character or projection. A third common factor was one whose highest

loading was on Wiener's subtle *Pa* or *PaS* scale (.85, .71, .76) (Table E). The whole *Pa* scale showed lower (though still substantial) loadings (.52, .34, .43) while Wiener's obvious paranoid subscale was unrelated to this factor (.06, −.10, .05). The designation "paranoid character" was applied to this dimension. This factor resembles Factor IV of Lingoes (1960). Paranoid character, without psychosis, is an accepted clinical concept and is an official medical diagnosis (American Psychiatric Association, 1952).

Factor 4: Conversion. A fourth factor that was extracted separately from Groups M, W, and C, had heavy loadings on *Hs* (with and without *K* correction) (.77, .61, .59; .80, .69, .65 corrected) and on *Hy* (.44, .72, .50), and much lower loadings on *D* (.22, .20, .10) (Table F). It was clearly conversion reaction, a long accepted clinical entity, and an official medical diagnosis (American Psychiatric Association, 1952).

The first three character types found as factors, the compulsive, the hysterical, and the paranoid, have been defined by psychoanalysts (Abraham, 1953) in terms of bodily zones of libidinal fixation. Others, however, (Freud, 1946; Freud, 1936) have defined these characters in terms of the predominant defense mechanism of the ego. These characters are marked, respectively, by the defenses of reaction formation, repression, and projection. Thus, the four factors most clearly cross-validated in the present study seem to be equivalent to four defenses: reaction formation, repression, projection, and conversion.

Factor 5: Oral aggression and delinquency. A fifth cross-validated factor was marked by high loadings on *Pd* (.63, .53, .58; or .67, .62, .65 with *K* correction) and on the Gough-Petersen delinquency or socialization scale (.34, .47, .52) (Table G). It was tentatively identified as oral aggression. It is less clearly identified than other factors in terms of psychodynamics and motivation. Gough (1957) described high scorers on *De* as "demanding," "rebellious," and "exhibitionistic." Cattell (1957b) used the same three adjectives in describing one pole of his Factor D, a factor which he omitted from the 16 PF for lack of an adequate measuring scale. It seems likely that Gough's delinquency-socialization is identical with Cattell's Factor D and with the factor found in the present study.

An experimental scale on rebellion, however, did not correlate in the expected manner with this factor (.06, −.21, −.03) (Table G). In Group M, the delinquent pole of the factor is correlated .22 with stubbornness, while among the women it is correlated .23 with submissiveness. Only further study on a new group can tell whether this is a genuine sex difference, or an accidental error of sampling. It fits in, however, with the general impression that women's delinquencies tend to be more passive and less aggressive than men's.

The tentative suggestion is offered that demandingness rather than rebellion may be the central feature of this factor. If so, it can be equated with the psychoanalytic concept of "oral aggression." This factor is not interpreted as socialization, conscience, or ego-ideal because of its insignificant loadings on Gough's responsibility scale (−.02, −.23, −.04).

Factor 6. A sixth factor, one of obsessive worrying, may be common to the three groups analyzed. Each group showed a factor positively loaded with *K* corrected *Pt* (.34, .40, .52), and less with uncorrected *Pt* (.15, .29, .37). The higher loadings on corrected than uncorrected *Pt* suggest (for reasons discussed above) that this factor is a personality variable rather than a response set. Groups M, W, and C also agreed in loading *Sc*-minus-*Pt* (−.24, −.34, −.39) and *Hy*-minus-*PtK* (−.67, −.50, −.16) on the same factor. The factor patterns from Groups M and W correspond well enough with the accepted (American Psychiatric Association, 1952) clinical concept of obsessive-compulsive pschyoneurosis. But the factor from Group C bearing nearest resemblance has more the flavor of general neuroticism, with additional loadings on *Hy* (.50), *Hs* (.33), *HsK* (.38) and *D* (.36).

Oblique Factors Less Clearly Established

There were several factors common to two of the three factor analyses. Group M and Group C showed a psychotic factor (Table I) with highest loadings on *Sc*-minus-*Pt*. It seems plausible that the absence of this factor in Group W was an accident of sampling.

A factor of sadism, cruelty, narrowmindedness, and prejudice appeared in Groups W and C (Table J). It was marked by loadings on the Berkeley authoritarian F Scale (abbreviated form), anti-Catholic prejudice, and an experimental scale designed to measure sadism.

Groups M and C shared a factor marked by high loadings on *Ma* (Table K). It was not significantly related to Welsh's *R*.

All three groups had factors related to masculinity-femininity (Table L), though there was no single variable with high loadings on the factors from all three groups. The variable *Mf*, which seems to do so, is artifactual in that it is scored differently for men and for women. Gough's *Fe* scale had rather high loadings on the masculinity factors from Groups M and C and lower loadings on the factor derived from Group W. In Group M, the masculine pole included inner-direction, and the feminine pole, other-direction, perhaps reflecting a difference between farmers and salesmen, respectively, in this Illinois sample. In any event, there is no theoretical reason for expecting "femininity" to mean the same in men as in women.

Both Group M and Group W showed factors whose highest loading was on the single item "I have had sex relations with at least three different people in my life." There is no similarity in other respects between the factor obtained from the men and that from the women (Table M). In women, high degree of sex experience was associated with measures of warmth, of delinquency, and of accepted passivity.[4]

Three other factors appeared in Group M alone. One was positively loaded with optimistic dependency, warmth, and phobia, and negatively loaded with *L*, *Si*, responsibility, and feminine masochism. Another consisted of (affirmatively) stubbornness, rebellion, dominance, and inner-direction, and (negatively) submission and *L*. The last comprised only delinquency affirmatively, and femininity, *R*, and stinginess in the negative direction.

Group W also showed three additional factors. One was loaded affirmatively with accepted passivity, sex experience, obsession, submission, and feminine masochism; and negatively with *Sc*-minus-*Pt*, sadism, rebellion, *ScK*, and *MaK*. It may be compared with Cattell's Factor E (1957a). Another comprised positive loadings on warmth, *PaS*, anti-Catholic prejudice, capacity for status, and phobia; and negative ones on *L*, inner-direction, *R*, *HyS*, and responsibility. The other had positive loadings on responsibility, sex experience, *L*, *Pa*, intellectual efficiency, and stinginess; and negative on phobia, femininity, and anti-Catholic prejudice.

Two other factors appeared in the combined group. One was a general hostility factor, loaded with H_2, *PaS*, H_1, stubbornness, *Pa*, and dominance. The other consisted of self-confidence and ambition, with loadings in one direction on capacity for status, intellectual efficiency, dominance, accepted passivity, responsibility, and warmth, and in the other direction for *R*, social introversion, and embarrassment.

Nature of the Oblique Factors

Some of the factors have been seen to correspond with diagnostic categories (compulsive personality, paranoid personality). But a basic conceptual difference must be pointed out. Diagnostic categories are discrete entities: a person either is or is not within the category, and the usual practice is to give a patient not more than one psychiatric label.

Factorial dimensions, on the other hand, are continuous. The question is not whether a person is or is not a paranoid character, but rather, how much paranoid character does he have? Further, a person may be high in two or more dimensions. A person high in both reaction formation and projection may, with equal justification, be called a compulsive personality or a paranoid one. Stagner and Moffitt (1956) have shown that persons high on a given personality trait do not show more

[4] Gough (personal communication, 1960) found, in 100 military officers, that those high on the Strong vocational interest keys for "banker" and for "mortician" reported early ages of first sexual intercourse, and were described as coarse, noisy, and masculine. It is to be predicted that this group should score low on the factor of repression or hysterical character, and at the "harria" pole of Cattell's "premsia versus harria" factor. While the evidence of the present study is ambiguous, it is consistent with clinical impressions that hysterical character has a complex effect on sexual behavior, delaying age of first sex experience, at least in males, yet facilitating later unconsciously motivated promiscuity, "hysterical acting out" (Table D, in women).

than chance resemblance in other traits. These considerations make it doubtful that psychiatric diagnoses are comparable to medical ones. Choice of diagnostic label seems not to be a fruitful point of dispute.

In almost every instance, *K* corrected scales proved to be factorially purer than the same scales without *K*. This finding reinforces Fricke's (1956) recommendation that *K* corrected scores be used in factor analysis. The increased factorial purity of *K* corrected scales appeared only after rotation to simple structure, and was not found in the centroid factors before rotation. Since *K* corrected scores are more valid (Meehl & Hathaway, 1946), this may be evidence that rotation to oblique simple structure produces factors more genuine than the unrotated ones.

The extraction of factors that have a clear clinical and psychodynamic meaning is a satisfying result. It provides some rapprochement between psychoanalytic and experimental approaches. Even more satisfying is the fact that several of the factors correspond to defenses of the ego. For of all psychoanalytic concepts, defenses are most amenable to objective study—far more so than libidinal fixation, for example. Defenses have effects in overt behavior and in perception, effects that can be studied through experiments on observable behavior (and to some extent have been, e.g., Eriksen, 1952, 1954; Mowrer—in rats—1940). Measurements of defenses by questionnaire scales derived from factorial studies could be useful in such behavioral studies. The clinical usefulness of such measurements also appears promising.

Summary

A Thurstone centroid factor analysis was done on 59 variables, including the original and some newer MMPI scales, on a group of 50 men, a group of 50 women, and the combined group of 100. Twelve or 13 factors were found in each group. The first 2 factors in all three groups resembled those previously found by Welsh and named *A* and *R*. These 2 factors were related to response sets, respectively, willingness to admit sickness and naysaying.

After rotation to oblique simple structure, an entirely different set of factors appeared. About six of these were replicated, being virtually identical in the three samples. Four of these corresponded to ego defenses of reaction formation, repression, projection, and conversion. Alternatively, the first three could be named anal or compulsive character, hysterical character, and paranoid character. The highest loadings on these factors, among previously published scales, were, respectively, Gough's rigidity-flexibility, Wiener's *Hy*-subtle, and Wiener's *Pa*-subtle. New experimental scales had higher loadings on some factors.

The anal character factor also confirmed Freud's concept, having substantial loadings on three experimental scales for orderliness, stinginess, and stubbornness.

A fifth common factor, loaded with *Pd* and with Gough's delinquency-socialization, was tentatively identified with oral aggression. A sixth factor was marked by *Pt*, and named obsessive worrying, but in the combined sample bore more resemblance to a factor of general neuroticism.

References

Abraham, K. A short study of the development of the libido, viewed in the light of mental disorders. (Originally printed 1924) In, *Selected papers on psychoanalysis*. New York: Basic Books, 1953. Ch. 26.

Adorno, T. W., Frenkel-Brunswik, E., Levinson, D. J., & Sanford, R. N. *The authoritarian personality*. New York: Harper, 1950.

American Psychiatric Association, Committee on Nomenclature and Statistics. *Diagnostic and statistical manual of mental disorders*. Washington, D. C.: APA Mental Hospital Service, 1952.

Barron, F. X. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, **17**, 327–333.

Cattell, R. B. *Factor analysis*. New York: Harper, 1952.

Cattell, R. B. *Handbook for the Sixteen Personality Factor Questionnaire*. Champa'gn, Ill.: Institute for Personality and Ability Testing, 1957. (a)

Cattell, R. B. *Personality and motivation structure and measurement*. Yonkers, N. Y.: World Book, 1957. (b)

Cook, E. B., & Wherry, R. J. A factor analysis of MMPI and aptitude test data. *J. appl. Psychol.*, 1950, **34**, 260–265.

Cottle, W. C. A factorial study of the multiphasic, Strong, Kuder, and Bell inventories using a population of adult males. *Psychometrika*, 1950, **15**, 25–47.

Couch, A. S., & Keniston, K. Yeasayers and naysayers: Agreeing response set as a personality variable. *J. abnorm. soc. Psychol.*, 1960, **60**, 151–174.

DeSoto, C. B., & Kuethe, J. L. The set to claim undesirable symptoms in personality inventories. *J. consult. Psychol.*, 1959, **23**, 496–500.

Dickman, K. W. *Oblimax rotation of factors.* Urbana, Ill.: University of Illinois, Digital Computer Laboratory, 1959. (Mimeo)

Drake, L. E. A social I.E. scale for the MMPI. *J. appl. Psychol.*, 1946, **30**, 51–54.

Edwards, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.

Eriksen, C. W. Individual differences in defensive forgetting. *J. exp. Psychol.*, 1952, **44**, 442–446.

Eriksen, C. W. Some personality correlates of stimulus generalization under stress. *J. abnorm. soc. Psychol.*, 1954, **49**, 561–565.

Fenichel, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.

Freud, A. *The ego and the mechanism of defense.* New York: International Univer. Press, 1946.

Freud, S. *The problem of anxiety.* New York: Psychoanalytic Quarterly Press, 1936.

Freud, S. Character and anal erotism. (Originally printed 1908) In, *Collected Papers.* Vol. 2. London: Hogarth, 1950. Ch. 4.

Fricke, B. G. Conversion hysterics and the MMPI. *J. clin. Psychol.*, 1956, **12**, 322–326.

Gough, H. G. The *F* minus *K* dissimulation index for the MMPI. *J. consult. Psychol.*, 1950, **14**, 408–413.

Gough, H. G. *Manual for the California Psychological Inventory.* Palo Alto, Calif.: Consulting Psychologists Press, 1957.

Gough, H. G., & Peterson, D. R. The identification and measurement of predispositional factors in crime and delinquency. *J. consult. Psychol.*, 1952, **16**, 207–212.

Hathaway, S. R. *The Minnesota Multiphasic Personality Inventory manual.* (Rev. ed.). New York: Psychological Corporation, 1951.

Hecht, S. An investigation into the psychology of masochism. Unpublished doctoral thesis, University of California, Berkeley, 1952.

Jones, E. Anal-erotic character traits. In, *Papers on psychoanalysis.* New York: Wood, 1913.

Karson, S., & Pool, K. B. The construct validity of the Sixteen Personality Factors Test. *J. clin. Psychol.*, 1957, **13**, 245–252.

Kendall, E. The validity of Taylor's manifest anxiety scale. *J. consult. Psychol.*, 1954, **18**, 429–432.

Lauterbach, C. G. The Taylor scale and clinical measures of anxiety. *J. consult. Psychol.*, 1958, **22**, 314.

Lazarus, R. S., Eriksen, C. W., & Fonda, C. P. Personality dynamics and auditory perceptual recognition. *J. Pers.*, 1951, **19**, 471–482.

Lingoes, J. C. MMPI factors of the Harris and the Wiener subscales. *J. consult. Psychol.*, 1960, **24**, 74–83.

Little, K., & Fisher, J. Two new experimental scales of the MMPI. *J. consult. Psychol.*, 1958, **22**, 305–306.

Luchins, A. R. Mechanization in problem solving: The effect of *Einstellung. Psychol. Monogr.*, 1942, **54**(6, Whole No. 248).

McKinley, J. C., & Hathaway, S. R. The Minnesota Multiphasic Personality Inventory: V. Hysteria, hypomania, and psychopathic deviate. *J. appl. Psychol.*, 1944, **28**, 153–174.

McNemar, Q. *Psychological statistics.* New York, Wiley, 1949.

Meehl, P. E., & Hathaway, S. R. The *K* factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *J. appl. Psychol.*, 1946, **30**, 525–564.

Mowrer, O. H. An experimental analogue of "regression" with incidental observations on "reaction-formation." *J. abnorm. soc. Psychol.*, 1940, **35**, 56–87. (Reprinted: In, *Learning theory and personality dynamics.* New York: Ronald, 1950. Ch. 13.)

Navran, L. A rationally derived MMPI scale to measure dependence. Unpublished doctoral dissertation, Stanford University, 1952.

Pinzka, C., & Saunders, D. R. Analytic rotation to simple structure: II. Extension to an oblique solution. *ETS res. Bull.*, 1954, No. 54–31.

Riesman, D., Glazer, N., & Denney, R. *The lonely crowd.* New Haven: Yale Univer. Press, 1950.

Rokeach, M. Generalized mental rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.*, 1948, **43**, 259–278.

Rosenzweig, S. The picture-association method and its application in a study of reactions to frustration. *J. Pers.*, 1945, **14**, 3–23.

Sears, R. R. *Survey of objective studies of psychoanalytic concepts.* New York: Social Science Research Council, 1942.

Stagner, R., & Moffitt, J. W. A statistical study of Freud's theory of personality types. *J. clin. Psychol.*, 1956, **12**, 72–74.

Taylor, J. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

Welsh, G. S. Factor Dimensions *A* and *R*. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine.* Minneapolis: Univer. Minnesota Press, 1956.

Wheeler, W. M., Little, K. B., & Lehner, G. F. J. The internal structure of the MMPI. *J. consult. Psychol.*, 1951, **15**, 134–141.

Wiener, D. N. Subtle and obvious keys for the Minnesota Multiphasic Personality Inventory. *J. consult. Psychol.*, 1948, **12**, 164–170.

Wiggins, J. S. Interrelationships among MMPI measures of dissimulation under standard and social desirability instructions. *J. consult. Psychol.*, 1959, **23**, 419–427.

Wiggins, J. S., & Rumrill, C. Social desirability in the MMPI and Welsh's factor scales *A* and *R*. *J. consult. Psychol.*, 1959, **23**, 100–106.

Williams, H. L., & Lawrence, J. F. Comparison of the Rorschach and MMPI by means of factor analysis. *J. consult. Psychol.*, 1954, **18**, 193–197.

(Received June 23, 1960)

Journal of Consulting Psychology
1961, Vol. 25, No. 4, 337–340

PLAY THERAPY LIMITS AND THEORETICAL ORIENTATION

HAIM G. GINOTT [1] AND DELL LEBO [2]
Child Guidance Clinic, Jacksonville, Florida

Child therapists agree that reasonable and consistent limits are necessary in play therapy with disturbed children. However, there are differences of opinion about the specific limits required in therapy. Some therapists (see Dorfman, 1951, p. 262) only set limits on activities that interfere with their ability to remain accepting of the child. They may allow a child to take toys home, break equipment, urinate on the floor, leave the playroom at will, or terminate treatment. Other therapists (Ginott, 1959) set limits on such activities. They only allow children to express themselves symbolically through words, play, and nonverbal gestures.

Judging from the literature, psychoanalytic therapists are less preoccupied with the problem of limits than are nondirective therapists. Extensive discussion of limits can be found in many nondirective publications (Axline, 1947; Bixler, 1949; Dorfman, 1951; Moustakas, 1953); only passing references to this subject appear in psychoanalytic writings (Schiffer, 1952; Slavson, 1952). On this basis, one might assume that psychoanalytic therapists employ fewer limits than nondirective therapists.

The authors know of no published study of the use of limits by therapists of different schools. The purpose of the present investigation was to discover whether limit setting is related to theoretical orientation. The experimental hypotheses were: (*a*) Therapists of different schools will not differ in the number of limits that they employ in play therapy. (*b*) Therapists of different schools will not differ in the kind of limits that they employ in play therapy.

[1] Now at New York University.

[2] Our thanks are expressed to Georgia Dreger and Dorothy Cohen for their assistance in compiling and checking the large array of data, and to Arthur Orgel for helpful advice.

METHOD

A questionnaire containing 54 discrete limits [3] was sent to 425 child guidance clinics and other agencies that treat children. The respondents were asked to identify themselves as being primarily nondirective, psychoanalytical, or "other," and to indicate (Yes, No, or Sometimes) whether they used a particular limit with children who were neither psychotic nor organic and between the ages of 3–10 years.

Questionnaires were returned by 227 play therapists; [4] of these 100 considered themselves to be psychoanalytic, 41 nondirective, and 86 of "other" schools. The number of Yes, No, and Sometimes responses to each of the 54 limits was totaled for each therapeutic school.

RESULTS

The mean number of limits used "ordinarily" and "sometimes" by therapists of the three schools are reported in Table 1. These means do not differ significantly; hence the results support the hypothesis that therapists of different orientations do not differ in the number of limits that they employ in play therapy.

[3] Copies of the questionnaire may be obtained from either of the writers.

[4] Most of the respondents were psychologists, a few were social workers and psychiatrists.

TABLE 1

MEAN NUMBER OF LIMITS "ORDINARILY" AND "SOMETIMES" USED BY PLAY THERAPISTS OF THREE SCHOOLS

Limit	Psychoanalytic	Nondirective	"Other"
Ordinarily	25	26	27
Sometimes	12	9	11

TABLE 2

SIGNIFICANTLY DIFFERENT PERCENTAGES OF LIMITS USED BY PLAY THERAPISTS OF THREE SCHOOLS

Limit	Use	School			χ^2
		Psychoanalytic	Nondirective	"Other"	
Enter playroom	No	35	54	45	11.064
	Yes	26	15	28	
	Some	39	31	24	
Pour water	No	43	54	30	9.965
	Yes	24	24	33	
	Some	18	12	16	
Paint cheap toys	No	58	78	57	19.593
	Yes	14	15	23	
	Some	28	7	19	
Paint costly toys	No	20	27	15	9.821
	Yes	53	61	54	
	Some	23	10	23	
Bring drinks, food	No	56	51	44	18.448
	Yes	4	20	23	
	Some	40	24	31	
Light matches	No	30	42	26	10.162
	Yes	30	34	36	
	Some	28	22	38	
Read books	No	54	71	48	18.321
	Yes	6	10	17	
	Some	40	19	35	
Do school work	No	35	54	22	23.668
	Yes	25	24	33	
	Some	40	22	44	
Break costly toys	No	9	19	6	10.965
	Yes	69	71	74	
	Some	21	10	16	
Hit therapist mildly	No	30	39	24	9.795
	Yes	42	42	54	
	Some	28	19	22	
Tie therapist	No	21	37	35	24.763
	Yes	44	54	44	
	Some	34	7	19	
Shoot therapist	No	16	31	15	12.004
	Yes	69	62	67	
	Some	15	7	16	
Fondle therapist	No	6	2	6	13.401
	Yes	62	83	65	
	Some	29	12	27	
Urinate or defecate	No	3	10	4	14.287
	Yes	75	71	86	
	Some	22	19	7	

Note.—Since a small number of items were not answered some columns do not total 100%.

To test the second hypothesis regarding the kinds of limits employed by different therapeutic schools, all responses to each of the 54 items were summed according to therapeutic orientation. These were converted into percentages and distributed into 3×3 chi squares. Of the 54 items, 14 achieved significance at better than the .05 level of confidence.[5] (On the basis of chance alone no more than 3 items might have achieved such significance.) The statistically significant items are shown in Table 2. These results do not support the second hypothesis that therapists of different orientations use the very same kind of limits in play therapy.

Discussion

Within the confines of the present study it is clear that therapists of the three approaches employ a similar number of limits in their work with children. While there are differences in the kinds of limits used by the three approaches, a considerable body of prohibitions are employed by all.

Thus, in the area of physical aggression against the therapist, practitioners of all schools concur to the same degree in prohibiting a child from squirting water on the therapist, painting his clothing, throwing sand, or forcefully attacking him. They differ, however, in that nondirectivists were significantly more permissive in allowing a child to shoot darts at or to hit the therapist and significantly less permissive in allowing a child to tie the therapist.

In the area of physical aggression against equipment, practitioners of all schools concur to the same degree in prohibiting a child from spilling sand, painting walls and furniture, starting fires, breaking windows and inexpentive toys, and throwing objects around the room. They differed, however, in that nondirectivists were significantly more permissive in allowing a child to pour much water into the sand box, paint and break expensive toys. The "other" therapists were the least permissive in allowing the painting of inexpensive toys.

[5] Three items whose statistical significance depended upon the Sometimes entries were omitted.

In the area of socially unacceptable behavior, practitioners of all schools similarly prohibited a child from smoking, using racial slurs, speaking or writing profanities in the playroom, making obscene objects, painting his face or clothing, undressing, and masturbating. They differed, however, in that the "other" therapists were significantly less permissive in allowing a child to urinate or defecate on the floor.

In the area of safety and health, practitioners of all schools similarly prohibited a child exploding a whole roll of caps, climbing on high window sills, drinking dirty water, or eating mud, chalk, or fingerpaints. They differed, however, in that the nondirectivists were significantly more permissive in allowing a child to light matches in the playroom.

In the area of playroom routines, practitioners of all schools similarly prohibited a child from taking home toys or clay objects, turning off the lights, leaving or overstaying, bringing in a friend, and talking to passers-by. They differed, however, in that the nondirectivists were significantly more permissive in allowing a child to decide whether or not to enter the playroom, to read books, and to do his school work there. The psychoanalytic therapists were significantly more permissive than members of the other two approaches in allowing a child to bring drinks and food into the playroom.

In the area of physical affection, practitioners of all schools similarly prohibited a child from sitting on their laps, hugging, and kissing them. They differed, however, in that nondirectivists were significantly less permissive in allowing a child to fondle them.

While 14 limits were used differently by the therapists of the present sample, the fact remains that practictioners of all schools concurred in the use of a large number of limits.

Summary

This study aimed to discover whether limit setting in play therapy is related to the therapist's theoretical orientation. Responses to a 54-item questionnaire on limits have been received from 100 psychoanalytic, 41 nondirective, and 86 "other" play therapists.

The results indicated that therapists of varied orientations did not differ in the number of limits used. While some significant differences were found in the kind of limits used, a considerable body of limits was employed by all.

REFERENCES

AXLINE, VIRGINIA M. *Play therapy*. Boston: Houghton Mifflin, 1947.

BIXLER, R. E. Limits are therapy. *J. consult. Psychol.*, 1949, **13**, 1–11.

DORFMAN, ELAINE. Play therapy. In C. R. Rogers, *Client-centered therapy*. Boston: Houghton Mifflin, 1951. Pp. 235–277.

GINOTT, H. G. The theory and practice of "therapeutic intervention" in child treatment. *J. consult. Psychol.* 1959, **23**, 160–166.

MOUSTAKAS, E. C. *Children in play therapy*. New York, McGraw-Hill, 1953.

SCHIFFER, M. Permissiveness versus sanction in activity group therapy. *Int. J. group Psychother.*, 1952, **2**, 255–261.

SLAVSON, S. R. *Child psychotherapy*. New York: Columbia Univer. Press, 1952.

(Received July 8, 1960)

Journal of Consulting Psychology
1961, Vol. 25, No. 4, 341-344

THE CONSTRUCT VALIDITY OF THE EDWARDS PPS HETEROSEXUALITY SCALE

E. JERRY PHARES AND CALVIN K. ADAMS
Kansas State University

The Edwards Personal Preference Schedule (PPS) (1954) appears to be increasingly used both in diagnosis and research. It purports to measure 15 personality needs growing out of H. A. Murray's work. The outstanding characteristics of the PPS appear to be a forced-choice item format, an attempt to control for social desirability in item choice, and 15 needs which are not obviously or necessarily connected with clinical pathology.

At the present time, however, many of the 15 scales of the PPS have not been systematically investigated. An exception would be the Achievement scale. Several studies have investigated the relationship between the PPS Achievement scale and the McClelland n Ach measure (Bendig, 1957; Himelstein, Eshenbach, & Carp, 1958; McClelland, 1958; Marlowe, 1959; Melikian, 1958). Still other studies have investigated the construct validity of some of the scales. Bernardin and Jessor (1957) utilized the Autonomy and Deference scales of the PPS as a measure of dependency and generally confirmed the construct validity of those two scales. Gisvold (1958) using an Asch group situation measure of conformity behavior found a significant negative relationship between it and the PPS Autonomy scale. No relationship between PPS Deference and conformity behavior was found, however. Zuckerman and Grosz (1958) used the Sway Test (a measure of suggestibility) and found low swayers scored significantly higher on the Autonomy scale than high swayers. The Deference and Succorance scales did not differentiate the groups. Zuckerman (1958) defining rebelliousness on the basis of sociometric scores could find only partial support for the hypothesis that rebellious subjects would score high on the Autonomy, Dominance, and Aggresion scales of the PPS. In another study of PPS n Ach, Worell (1960) demonstrated that subjects high on this scale showed significant superiority over low subjects in two verbal learning situations. In a factorial investigation of the entire PPS, Levonian, Comrey, Levy, and Proctor (1959) found a discrepancy between what the PPS is designed to measure and its actual factorial item content.

The present study is an attempt at the construct validation of the Heterosexuality scale of the PPS as regards males. This variable is described by Edwards (1954) as follows:

> To go out with members of the opposite sex, to engage in social activities with the opposite sex, to be in love with members of the opposite sex, to kiss those of the opposite sex, to be regarded as physically attractive by those of the opposite sex, to participate in discussions about sex, to read books and plays involving sex, to listen to or tell jokes involving sex, to become sexually excited (p. 5).

Therefore, given this definition of heterosexuality how might males who score high on this scale behave differently from males who score low? Or, stated another way, will behavior in experimental situations developed in the light of the characteristics of such a heterosexuality construct relate to behavior on the Heterosexuality scale of the PPS? If so, the construct validity of the scale will have, in part at least, been supported.

Selection of subjects for this study was based upon the administration of a scale to 170 males in two general psychology classes at Kansas State University. The scale consisted of the 28 items of the PPS Heterosexuality scale plus 22 buffer items selected at random from other PPS scales.[1] From

[1] There is some evidence that removing items from the context of a standardized test may alter the nature of the items and responses to them (Edwards, Wright, & Lunneborg, 1959).

this procedure a high heterosexual group was drawn composed of 20 subjects with scores between 16 and 28. A low group contained 20 subjects with PPS scores between 1 and 8. The same two groups of subjects performed in both phases of the study to be reported below.

Phase 1

Relationships between needs and esthetic preferences would appear logical ones given the pervasive role needs are assumed to play. Such a relationship between n Achievement and esthetic preferences has been investigated by Knapp (1958). For our purposes, if high PPS heterosexual subjects find sex and sex related activities more congenial and pleasurable than do low PPS subjects, it seems reasonable to conclude that their esthetic preferences for photographs might reflect this. Therefore, it was predicted that high PPS heterosexual subjects would place a higher esthetic value on photographs involving sexual elements in varying degrees than would low PPS heterosexual subjects.

Procedure

Sixty black and white photographs were selected principally from various art and photography magazines, but also from other sources. On a common sense basis, the photographs were classified as either sexual or nonsexual. The former consisted of such subjects as female nudes, female facial portraits, "cheesecake," boy and girl holding hands or kissing, etc. The nonsexual photographs ranged from skylines of New York to street beggars, animals, children, etc. Each photograph was then examined by three judges who independently determined its sexual or nonsexual character. Using unanimous agreement as a criterion eight photographs were eliminated as ambiguous. Next, 47 unselected males were drawn from general psychology classes at Kansas State University. In three group sessions of about 15 subjects each, the remaining 52 photographs were consecutively projected onto a screen. The subjects were asked to rate each photograph on a five-point scale along a dimension of esthetic value. This enabled a value to be assigned to each photograph obtained by summing the ratings of the subjects and dividing by 47.

The remainder of the pre-experimental work consisted of arranging the above photographs into seven groups of six photographs each. Each group of six contained four nonsexual and two sexual photographs, all of approximately equal value. This procedure eliminated 10 photographs whose values were either too high or too low to easily fit into any of the seven groupings. Following this, each group of six photographs was pasted onto $21'' \times 28''$ white cards in two rows of three each. The positions of the sex photographs were randomly distributed throughout the series.

In the experiment itself, subjects in both groups were told they were participants in a study to develop an art appreciation test and were asked to to rank the photographs in each group from one to six in terms of how artistically pleasing they were. Subjects were allowed a maximum of $1\frac{1}{2}$ minutes to examine each group of six, whereupon they made their rankings and the next group of six was presented. Each subject was given a score which consisted of the sum of his rankings of the 14 sex pictures.

Results

The mean and standard deviation of the high heterosexual group were 45.2 and 12.8, respectively, while the corresponding mean and sigma for the low group were 57.6 and 9.1.[2] This mean difference of 12.4 with a *SE* of the difference between means of 3.5 yields a t value of 3.5, significant at the .001 level. It is interesting to note also that only two cases in the high group exceeded the median of the low group and only one case exceeded the ninetieth percentile of the low group.

From these data it is clear that subjects high on the PPS Heterosexual scale place a higher value on photographs involving sexual elements than do subjects low on that scale.

Phase 2

In this study it was hypothesized that subjects high on the PPS Heterosexuality scale would show greater retention of material encouraging the importance of sexual information, dating, and sex education in the prevention of mental illness than would low PPS Heterosexuality subjects. This prediction was based on the assumption that subjects should better learn and retain material supportive of their needs than material opposed to them or that contravaluant material should be more disruptive of the retention process than supportive material. This prediction would also be consistent with research on attitudes and their role in learning and retention (Edwards, 1941; Levine & Murphy, 1943).

Procedure

As a warmup, subjects were given a brief and easy digit span test. Subjects were then read two short passages. Following each, they were given 4 minutes to write down as much of it as they could remember.

[2] The higher the score, the lower the esthetic value.

The first or neutral passage was Memory Selection (A) from the Wechsler Memory Scale (1945). The second selection was the sexual passage which read:

> Dr. Paul Rogers/ the eminent/ New/ London/ psychiatrist/ recently/ spoke/ on the importance of/ sex education/ and dating/ in the prevention/ of mental illness./ He urged/ males/ to learn/ much/ more about sex/ and sexual functions./ He also stressed/ the importance/ of active dating/ and experience with girls/ in promoting/ happiness./ The speech was/ very/ well received/ and increased/ his stature/ as an authority/ on sex./

Each subject's score was the total number of units recalled on the sexual passage subtracted from the total number recalled on the neutral passage. To eliminate negative numbers a constant of 10 was added to subjects' scores. This procedure enabled the control of such extraneous variables as intelligence and learning ability which might have accounted for differential retention in the two groups. Twelve protocols were randomly selected from the two groups of subjects to establish interjudge scoring reliability. For the Wechsler passage the interjudge scoring reliability coefficient for the two independent judges was .97, and for the sex passage, .92. The senior experimenter's scoring alone was used in testing the hypothesis.

Results

The mean retention score and standard deviation of the high heterosexual group were 10.95 and 3.8, respectively, and the corresponding mean and sigma for the low group were 12.0 and 2.8.[3] The difference between the means is thus 1.05 and the *SE* of the difference between means 1.06. A t test gave a value of .9906. This result, while not significant, is in the predicted direction.

Discussion

The hypothesis of the first study was confirmed in a highly significant manner. The second study produced nonsignificant results although in the expected direction. In view of the latter consideration, a t test was run again on the retention data after eliminating the five lowest scoring high heterosexual subjects and the five highest scoring low subjects. In this fashion, three subjects with PPS scores of 16, two with scores of 17, and five with scores of 8 were thrown out thus making more extreme groups of 15 each. The number of cases eliminated was determined arbitrarily without reference to any distribution characteristics. Nor were there any "holes" likely to inflate a trend. As it happened, this procedure entirely eliminated all subjects with scores of either 16, 17, or 8—there was no need to "choose" which subjects to include and which to exclude. This procedure resulted in mean retention scores of 10.0 and 12.2 for the high and low groups, respectively. The t value, based on a mean difference of 2.2 and a *SE* of the difference between means of 1.15, was now 1.9, significant at the .06 level. Matching the groups pair-wise on the basis of their scores on the neutral passage and assuming the discrepancy scores to be correlated, lowered the obtained p level to only .05. These latter findings suggest either that the Edwards Heterosexuality scale is somewhat nondiscriminating except at the extremes or that our retention measure was not as sensitive as it might have been. In retrospect it seems that we could have constructed a more threatening or need-engaging sex paragraph or perhaps a measure more easily capable of being scored for distortion.

As a test of whether the two criterion measures were related within subjects, subjects' scores on the retention task were correlated with their scores on the photograph task. The correlation was only .155. Again however, eliminating the aforementioned 10 subjects raised the r to .29. This coefficient is not significant, but in view of the increase may suggest a modicum of generality, especially assuming a more sensitive retention measure.

All in all, the strong evidence provided by the results of the first hypothesis and the suggestive evidence of the second hypothesis appear to give support to the construct validity of the Edwards PPS Heterosexuality scale as regards males and underscore particularly its potential usefulness for research purposes.

Summary

In this study the Edwards PPS Heterosexuality scale was investigated with respect to its construct validity. Groups of high and low scoring males were utilized. Two investigations, one involving esthetic preferences and the other retention scores, were carried out. In the first case it was found that high PPS males placed a significantly higher es-

[3] The higher the score, the lower the retention of sex material.

thetic value on sexual photographs than did low subjects. In the second study suggestive evidence was found that high subjects exhibit better retention of sexual material than do low subjects.

Generally, the evidence appears supportive of the construct validity of the Edwards PPS Heterosexuality scale.

REFERENCES

BENDIG, A. W. Manifest anxiety and projective and objective measures of need achievement. *J. consult. Psychol.*, 1957, **21**, 354.

BERNARDIN, A. C., & JESSOR, R. A construct validation of the Edwards Personal Preference Schedule with respect to dependency. *J. consult. Psychol.*, 1957, **21**, 63–67.

EDWARDS, A. L. Political frames of reference as a factor influencing recognition. *J. abnorm. soc. Psychol.*, 1941, **36**, 34–50.

EDWARDS, A. L. *Personal Preference Schedule*. New York: Psychological Corporation, 1954.

EDWARDS, A. L., WRIGHT, C. E., & LUNNEBORG, C. E. A note on "social desirability as a variable in the Edwards Personal Preference Schedule." *J. consult. Psychol.*, 1959, **23**, 558.

GISVOLD, D. A validity study of the autonomy and deference subscales of the EPPS. *J. consult. Psychol.*, 1958, **22**, 445–447.

HIMELSTEIN, P., ESCHENBACH, A. E., & CARP, A. Interrelationships among three measures of need achievement. *J. consult. Psychol.*, 1958, **22**, 451–452.

KNAPP, R. H. n achievement and aesthetic preference. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. New York: Van Nostrand, 1958.

LEVINE, J. M., & MURPHY, G. The learning and forgetting of controversial material. *J. abnorm. soc. Psychol.*, 1943, **38**, 507–515.

LEVONIAN, E., COMREY, A., LEVY, W., & PROCTOR, D. A statistical evaluation of Edwards Personal Preference Schedule. *J. appl. Psychol.*, 1959, **43**, 355–359.

MCCLELLAND, D. Methods of measuring human motivation. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. New York: Van Nostrand, 1958.

MARLOWE, D. Relationships among direct and indirect measures of the achievement motive and overt behavior. *J. consult. Psychol.*, 1959, **23**, 329–332.

MELIKIAN, L. H. The relationship between Edwards' and McClelland's measures of achievement motivation. *J. consult. Psychol.*, 1958, **22**, 296–298.

WECHSLER, D. A standardized memory scale for clinical use. *J. Psychol.*, 1945, **19**, 87–95.

WORELL, L. EPPS n achievement and verbal paired associates learning. *J. abnorm. soc. Psychol.*, 1960, **60**, 147–150.

ZUCKERMAN, M. The validity of the Edwards Personal Preference Schedule in the measurement of dependency-rebelliousness. *J. clin. Psychol.*, 1958, **14**, 379–382.

ZUCKERMAN, M., & GROSZ, H. J. Suggestibility and dependency. *J. consult. Psychol.*, 1958, **22**, 328.

(Received July 11, 1960)

Journal of Consulting Psychology
1961, Vol. 25, No. 4, 345–352

SECOND-ORDER PERSONALITY FACTOR STRUCTURE IN THE OBJECTIVE TEST REALM[1]

R. B. CATTELL
University of Illinois
R. R. KNAPP
United States Naval Personnel Research Activity, San Diego, California
AND I. H. SCHEIER
University of Illinois

A comprehensive, empirical system of personality factor dimensions based on factor analyses of a battery of objective tests has been described in previous publications (Cattell, 1955a; Cattell, 1957a). The initial emphasis in the development of this system has been to establish these as *first-order factors,* that is factors which are based on correlations between *tests,* and which therefore adhere closely to the detailed information available in tests. Recently, as first-order factor description has attained increasing replication, permitting more exact conceptualization and measurement, it has become possible to shift attention to *second-order factors,* that is factors based on the correlations between first-order factors (Cattell, 1956, 1957a). These correlations may be determined either (*a*) by measuring the factors by factor batteries, in which case they are attenuated by test unreliability and disturbed by invalidity, or (*b*) from the plots of rotations to simple structure in the first-order domain. In the latter case, there is no attenuation effect and relatively slight disturbance by invalidity, if and as the rotation is good and enough variables are employed to define the hyperplanes. Research needs to take both approaches, since the degree of consensus throws light on the magnitude of error. Both approaches are considered in this article.

Since factor analysis is a parsimonious operation, there will be fewer second-order than first-order factors, just as there are fewer first-order factors than there are tests (Cattell, 1952). Second-order factors may therefore be looked upon as relatively broad descriptive categories, interpretable as representing general organizing influences in personality. Due to the limited number of explanatory concepts which the human mind seems able to juggle at one time, as well as to the analytic defects of behavioral analysis at the level of premetric general and clinical observation, the concepts of most psychologists fit factors at the second-order rather than the first-order level. For example, the personality concept measured by the questionnaire second-order extraversion factor in the 16 PF (Cattell, 1956; Cattell, Saunders, & Stice, 1957) is probably referred to more often than are concepts measured by the first-order components: Cyclothymia, Surgency, Parmia (Cattell, 1957a; Cattell, Saunders, & Stice, 1957). Both levels have their use and, although secondaries (second-orders) lose some of the predictive power of primaries (first-orders), knowledge of the second-order structure is extremely important for understanding personality, developmentally and in action.

The position achieved by research to date, in respect to orders and media of observation, is as follows:

1. The establishment of first-order personality factors in the questionnaire (verbal, self-evaluative) and life record (behavior rated *in situ*) media of measurement (e.g.

[1] The opinions expressed are those of the writers and are not necessarily shared by the Department of the Navy.

Cattell, 1957a; Cattell, Saunders, & Stice, 1957).

2. The tentative establishment of first-order personality factors in the objective test (or T) realm (Cattell, 1957a; Cattell & Scheier, 1959; Scheier & Cattell, 1958)—where "objective" tests are understood as being tests which are relatively disguised in purpose, difficult to fake, and which are based on the subject's performance in miniature situational tests rather than self-report (Cattell, 1958; Scheier, 1958).

3. The establishment of second-order personality factors in the questionnaire (or Q) medium of measurement (Cattell, 1956).

4. The establishment of *some* matches across Q and T media. After separate factorization within each of the two media (in 1 through 3 above), factor scores are correlated to seek matches in the two series. Results in this area are not yet definitive, but they strongly indicate that (*a*) at least four second-order factors in questionnaires match (i.e., measure the same dimensions as) first-order factors in objective tests (Cattell & Scheier, 1959, 1961; Scheier & Cattell, 1958), and (*b*) even when the entire second-order questionnaire factor realm is accounted for and matched as best one can with objective test factors, all but five or six of the objective test factors lack substantial questionnaire association—i.e., they do *not* have questionnaire factor equivalents at either the first- or second-order Q level. A possible inference is that many objective test factors are getting at areas of personality which questionnaire factors do not, and probably never can, measure.

5. First explorations have been made of the second-order factors among the objective test medium primaries. As Paragraph 4 above indicates, these might be expected to represent broader influences than those in questionnaire and rating primaries. The present paper concerns itself with organizing the evidence from these recent studies.[2]

Method

The researches available for collation are five in number, four being based on method of analysis (*b*) and one on method (*a*) as described in the first paragraph of this paper. All studies operated upon individual difference patterns—i.e., they were not incremental or P technique studies concerned with *state* factors, but rather, R technique analyses concerned with trait factors (Cattell, 1952, 1957a). For closer comparability, not confusing population differences with experimental error, the studies considered in the present paper were all based on young male American adults. For ease of reference in subsequent tables and discussion, they are referred to by temporary symbols used in other publications from the senior author's laboratory (C_5, C_6, etc., below). These studies are described in necessary detail elsewhere (Cattell, 1955b; Cattell & Scheier, 1959, 1961; Scheier & Cattell, 1958), but the essential characteristics are as follows:

C_5. 500 United States Air Force males measured on 128 variables. Rotation to oblique simple structure at the first-order yielded 16 factors (Cattell, 1955b) and was followed by a second-order analysis (Method *b*) of the correlations among first-orders (see Appendix 12 in Cattell, 1957a).

C_6. 250 United States Air Force males measured on 64 variables. Rotation to oblique simple structure at the first-order yielded 15 factors (Cattell, 1955b) and was followed by a second-order analysis (Method *b*) of the correlations among first-orders (see Appendix 12 in Cattell, 1957a).

R_1, R_2. Two studies each of which measured the same 86 male college undergraduates, with 120 and 103 variables, respectively (69 of which were common to both studies). Rotation to oblique first-order simple structure, reported on elsewhere (Cattell & Scheier, 1959; Scheier & Cattell, 1958), yielded, respectively, 15 and 17 factors, and was followed by second-order analysis (Method *b*) of the correlations among first-order factors. These two second-order resolutions are published here for the first time.

N_1. 315 United States Navy male Submarine School candidates[3] were scored on 18 first-order objective test factors as measured by the O-A Battery (Cattell, 1955a). That is, in the N_1 study each person's score on each factor was obtained from a combination of test scores demonstrated, by consistent and substantial loadings on each factor, to give a best possible estimate of that factor. Correlations were computed among these battery *scores* (Method *a* above), and factors were then extracted and rotated to oblique simple structure. The second-order results are reported for the first time. The N_1 study is thus unique in that *factor score* correlations were used instead of the C_F matrix (Cattell, 1952) derived from

[2] Eventually, second-order objective test factors will have to have their relations checked with first- and second-order Q factors. Presumably, if the evidence of Paragraph 4a above holds up, no direct matches will be found because second-order objective test factors occur at a higher order than any known Q factors, namely at a *third*-order, relative to first-order Q factors.

[3] These data were collected at the United States Naval Medical Research Laboratory, New London, Connecticut, under Bureau of Medicine and Surgery Project NM 23 02 20.

oblique rotations, as was the case in the other four studies.

Second-order factorization in all of the five studies used tests of completeness of factor extraction described elsewhere (Cattell, 1952) and pursued simple structure blindly to a plateau at which hyperplane count proved unimprovable. At that point the factor patterns were inspected and a process of matching (correlation of loading patterns) was begun, factor by factor, with each of the other four studies. This matching revealed tolerably good convergence of results in the five studies and the matched factors and their loadings are therefore set out in Tables 1 through 7 below, each table representing a second-order objective test factor which we now believe to be reasonably well confirmed. All five studies replicate the same seven second-order factors, and although a given factor is often less clear in one series than in another, no alternative matching would be nearly as satisfactory as that presented.

The layout of these tables follows the usual conventions, as follows:

1. Each table is headed by a contingent verbal title for the second-order factor and a roman numeral. The verbal title may be modified as more information becomes available on the factor, but its roman numeral remains constant for easy identification in present and future studies. Each verbal title is bipolar, the positive (high score) pole being on top and and the negative in parentheses.

2. First-order factors—the "variables" at this order of analysis—are identified in the left-hand column of each table, (*a*) by their Universal Index or UI numbers (Cattell, 1957b), which are constant across all published researches, and (*b*) by their verbal titles.

3. The figures in Columns 3 through 7 are the loadings of first-orders on the second-order, each column being designated by the symbol for the research in which the results were found. These values diverge from the true values, in the case of C_5, C_6, R_1, and R_2, by reason of imperfect simple structure positions in first- and second-order relations, and in the case of N_1 by attenuation due to unreliability of the factor batteries. Inconsistencies of sign between one study and others are indicated by enclosing the atypical value in brackets. A positive sign on a loading means that the positive pole of that factor, as it is standardly written in various textbooks (Cattell, 1957a; Cattell & Scheier, 1961) goes positively with the second-order factor pole as written, and oppositely for a minus loading. Since, in reading titles on the left, readers like to see at once which "go together," we have followed the practice of writing in the pole which is consistent with the positive direction of the *second-order factor*. That is to say, when a loading in the numerical columns is negative (for the positive pole of the first-order factor), *we have already reversed the title to put the negative pole of the factor in this title column*. Therefore, the verbal title of the first-order already gives the direction in which (pole at which) it is related to the second-order factor. The signs on the loadings merely tell whether this is the positive (high score) or negative (low score) pole of the first-order, as it is usually scored and thought of.

4. The second column gives the average loading, which, by reason of the attenuation in N_1, would be expected to err systematically, slightly below the true value. The given rank order in listing the primaries is not literally the declining order of their mean loading on the second-order, but an estimate of the order of importance of the primaries, made by taking into account also the consistency and frequency of replication of the result.

5. The tables report data only on the first-orders which are most highly and consistently associated with second-orders. The full tables, showing the loading of *all* first-orders present in each study, including those with lesser to essentially zero relationship, are preserved in tables available from the American Documentation Institute.[4]

[4] Complete loading tables for all first-order factors have been deposited with the American Documentation Institute. Order Document No. 6758 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

TABLE 1

F(T)I: Tied Socialization or Superego (vs. Absence of Cultural Introjection)

First-Order Factor	Loadings in Studies					
	Average	C_5	C_6	R_1	R_2	N_1
UI 20 Comention, Conformity to Cultural Standards	+36	+62	+46	+03		+34
UI 28 Rigid Superego	+26	+64	+13		+22	+05
UI 1 Low General Intelligence	−34	−34	−33			
UI 19 Subduedness	−20	−30	−37	−09	−06	−20
UI 25 Accurate Realism	−33	−04		−52	−15	−62
UI 35 Self-Reliance	+44			+44		
UI 32 Extraversion	−19	−17	−38	−01	−21	−16

TABLE 2

F(T)II: Expansive Ego (vs. History of Difficulty in Problem Solving)

First-Order Factor	Loadings in Studies					
	Average	C_5	C_6	R_1	R_2	N_1
UI 16 Harric Assertiveness	+34	+47	+56	+32	+36	00
UI 1 High General Intelligence	+28	+42	+13			
UI 19 Promethean Will	+23	+18	+33	+21	+06	+37
UI 23 Ergic Regression	−21			−08	−06	−48
UI 36 Self-Sentiment Development	+29			+29		
UI 18 Naivete	−15	00	−25	−03	−26	−22

Data and Interpretation for Second-Order Factors

Factor F(T)I, Tied Socialization, has claims to being the largest second-order factor. There is high internal psychological consistency (with the possible exception of low intelligence) in terms of the characteristics of the first-order factors involved. A clinician might well claim it to be proof of a broad superego organization, and some solid support for this lies in the finding (Cattell & Scheier, 1961) that neurotics are significantly higher than normals on this second-order factor. However, in the label "Tied Socialization" we have favored an hypothesis which *may* specialize to a superego definition, but which is at present broader. It implies that this pattern represents the extent to which the individual has accepted the culture patterns and standards of the group. It has overtones of extraversion and realistic contact, subduedness, receptivity, and lower intelligence (uncritical?), but the central factors are UI 20, Comention (conventional acceptance of values) and UI 28, Rigidity of Superego, which we believe the contingent title aptly designates.

Except for UI 23(−), there is again considerable psychological consistency in F(T)II, here in the sense of willfulness ("will power") and a high development of the self-sentiment. Once again, too, the clinician might claim that this particular clustering of first-order factors gives support to an observational concept at the second-order level, in this case, the concept of a broad dynamic organization that is called "ego development" or "ego strength." The essential unifying concept, covering higher intelligence and the assertive characteristics, appears to be, if an environmental explanation is adopted, a history of success in managing ergic (drive) satisfactions. This interpretation would also fit psychoanalytic concepts of ego strength The

TABLE 3

F(T)III: Temperamental Ardor (vs. Temperamental Apathy)

First-Order Factor	Loadings in Studies					
	Average	C_5	C_6	R_1	R_2	N_1
UI 21 Exuberance, Energetic Spontaneity	+31	+70	+40	+23	+11	+13
UI 20 Comention, Conformity to Cultural Standards	+27	+48	+12	+08		+38
UI 1 Low General Intelligence	−28	−15	−41			
UI 19 Promethean Will	+21	+21	+12	+18	+54	+01
UI 27 Alert Control	−19				−29	−09
UI 32 Extraversion	−16	(10)	−34	−55	(03)	−06

TABLE 4

F(T)IV: HIGH EDUCATED SELF-CONSCIOUSNESS (VS. LOW EDUCATED SELF-CONSCIOUSNESS)

First-Order Factor	Loadings in Studies					
	Average	C_5	C_6	R_1	R_2	N_1
UI 22 Corticalertia, Cortical Alertness	+31	+64	+40	+08	(−02)	+47
UI 18 Shrewdness	+28	+66	+24	+40	+07	+01
UI 25 Imaginative Tension	+17	+45		+12	+02	+08
UI 36 Self-Sentiment Development	+51			+51		
UI 30 High General Reactivity	−16	−03	−08		(+05)	−57
UI 29 Low Adaptation Energy	−15	−01	−08		−53	(+04)
UI 33 Dourness	+20					+20

chief criterion evidence on this factor shows it distinguishing even more powerfully than F(T)I between neurotics and normals, the normals being significantly higher than neurotics in Expansive Ego (Cattell & Scheier, 1961).

A noteworthy feature of F(T)III is that it tends to pick out those factors which individually appear to have relatively high genetic determination (Cattell, 1957a). Previous multivariate experiments (Cattell, Stice, & Kristy, 1957) using groups of identical and fraternal twins and siblings, some of which were reared together and some apart, have attempted to assess the contribution which heredity and environment may make to certain of the objective test factors. In particular it appeared that heredity was the main determiner in F(T)III Factors UI 1, Intelligence, and UI 20, Comention. Hereditary and environmental influences appeared to be about equal in determining Factors UI 19, Promethean Will, and UI 27, Alert Control, which are found on F(T)III. Conceivably, these genetic determinations all arise from a single genetic influence covering all primaries contributing to this second-order factor. However, it is difficult to imagine in polygenic determination how a series of genes would happen to coincide in repeatedly affecting several things in the same way. This is a provocative finding for psychological genetics. Meanwhile, we shall label the second-order "Temperamental Ardor," since, except for UI 20, the central character is a willful exuberance and ardor of temper. Consistent with this interpretation, F(T)III has been found to be significantly higher for hospitalized neurotics of sociopathic type than it is for normals (Cattell & Scheier, 1961).

F(T)IV tends to be loaded by what appear to be largely environmentally determined factors. In previous investigations (Cattell, Stice, & Kristy, 1957) environment appeared to be the main determiner in F(T)IV Factors UI 22, Corticalertia, and UI 29, Low Adaptation Energy-vs.-Overresponsiveness. It probably has something to do with education in the sense of developing alert, shrewd, and imaginative qualities. It also involves much ex-

TABLE 5

F(T)V: HISTORY OF INHIBITING, RESTRAINING ENVIRONMENT (VS. LAXNESS)

First-Order Factor	Loadings in Studies					
	Average	C_5	C_6	R_1	R_2	N_1
UI 17 Inhibition	+36	+32	+54	+45	+03	+47
UI 23 High Mobilization of Resources	+18			+25	+19	+10
UI 31 Wary Realism	+16	+07	(−12)			+52

TABLE 6

F(T)VI: NARCISTIC DEVELOPMENT (VS. ENVIRONMENTAL CONTACT AND INVESTMENT)

First-Order Factor	Loadings in Studies					
	Average	C_5	C_6	R_1	R_2	N_1
UI 26 Narcistic Self-Will	+33	+59	+31	+50	+25	00
UI 27 Apathy-Fatigue	+30				+36	+23
UI 34 Autistic Nonconformity	+51			+51		

plicit self-awareness and "canniness." Possibly, an upbringing in a critical, competitive, high-standard home atmosphere could generate such a pattern. The above interpretation is, at least, not inconsistent with the fact that neurotics with marked sociopathic trends have significantly lower scores than normals on this factor (Cattell & Scheier, 1961).

Our hypothesis is that F(T)V represents a dimension of personality resulting from an environment in which considerable inhibition prevails. The somewhat unexpected role of UI 23, here as in F(T)II, suggests that there may be some need to modify our hypotheses about this primary. However, it is reasonable to interpret UI 23 as unused reserves of energy and it might then fit here as the cumulative result of inhibited, undischarged reactivity.

F(T)VI is a rather narrow factor which nevertheless has a consistent character of narcistic and autistic development. It is in some sense a false ego development, a moving out of contact with reality. This hypothesis is supported by its being significantly high in neurotics with sociopathic trends, but not in typical neurotics (Cattell & Scheier, 1961).

In F(T)VII, we may have a form of tension wider than Free Anxiety, UI 24, and therefore perhaps representing total drive tension as it is oriented and controlled toward achievement. On the other hand, an argument could be made from present inclusions and exclusions that the central influence in the factor is insecurities connected with the self-concept. We shall keep interpretive eventualities open by a descriptive title indicating anxiety and related drive tensions centered on insecurity, and under control of an achievement goal.

Now that the patterns of second-order objective test factors have been presented and tentatively interpreted, it is logical to ask if very broad *third*-order factors can be found. Accordingly, we computed the correlations among second-order T factors, in the three most recent studies (R_1, R_2, N_1) and averaged these values in Table 8. These correlations among second-orders are generally very low. Over half of the r's are .10 or less, and only one, between F(T)III and F(T)VI, could confidently be called statistically significant. It is possible that a third-order might be found, including the second-orders of Tem-

TABLE 7

F(T)VII: HIGH TENSION TO ACHIEVE, CONTROLLED DRIVE TENSION LEVEL (VS. LOW TENSION TO ACHIEVE)

First-Order Factor	Loadings in Studies					
	Average	C_5	C_6	R_1	R_2	N_1
UI 24 Free Anxiety	+40	+38	+38	+52	+42	+30
UI 18 Shrewdness	+23	+14	+38	+34	+18	+09
UI 30 High General Reactivity	−19	−56	−12		00	−07
UI 25 Imaginative Tension	+18	+49		+14	(−07)	+17
UI 19 Promethean Will	+16	+13	+10	+07	+05	+46

TABLE 8

CORRELATIONS AMONG SECOND-ORDER OBJECTIVE-TEST FACTORS

I	II	III	IV	V	VI	VII
	−.10	−.06	−.17	.19	−.18	.07
		.01	−.07	−.17	.07	.12
			−.18	−.14	.38	−.14
				−.08	−.06	.05
					−.08	−.07
						−.04

Note.—Average over R_1, R_2, and N_1 studies.

peramental Ardor—F(T)III—and Narcistic Development—F(T)VI; but such factorization should await further replication and confirmation of second-order results. However, even now, it is possible to make a general prediction that, for practical purposes, oblique factorizations do not permit an indefinitely large series of higher-and-higher order factorizations. The correlations definitely become lower as one moves to higher orders; that is, correlations among second-orders are generally lower than correlations among first-orders, and correlations among first-orders are generally lower than correlations between tests.

There is no logical reason why the same basic data should not be describable by using sets of categories, alternatively at lower or higher orders, with the broader categories being fewer but missing more detail. On the other hand, it would be somewhat confusing and discomfiting if this process were to prove empirically and practically possible over, say, five, six, or even a greater number of distinct levels of generality (orders). The data of Table 8 strongly suggest an upper limit of three orders for objective test data, and, since there are only four well-confirmed second-order factors in questionnaires (Cattell, 1956; Cattell & Scheier, 1961), it is almost certain that they too will yield no more than three orders.

SUMMARY

1. Correlations among 20 first-order personality factors, in objective tests, have been obtained in five separate studies employing young male adult samples.

2. An oblique simple structure factoring, independently on each of the five correlation matrices, yields seven factors, the general form of which is confirmed by matching across the five studies.

3. With relatively short supportive discussion, these factors have been indexed and named as follows:

F(T)I, Tied Socialization
F(T)II, Expansive Ego
F(T)III, Temperamental Ardor
F(T)IV, Educated Self-Consciousness
F(T)V, History of Inhibiting, Restraining Environment
F(T)VI, Narcistic Development
F(T)VII, Tension to Achieve

4. Space dictates restricting this presentation to evidence of the patterns, their matching, and a few criterion associations. Elsewhere, a continuation will be made of the fuller theoretical development required by this first demonstration of consistent personality structure at the more pervasive level of second-order objective test factors.

5. Possibility of third-order factors is briefly discussed.

REFERENCES

CATTELL, R. B. *Factor analysis: An introduction and manual for the psychologist and social scientist.* New York: Harper, 1952.

CATTELL, R. B. *Handbook for the objective-analytic personality test batteries.* Champaign, Ill.: Institute of Personality and Ability Testing, 1955. (a)

CATTELL, R. B. Psychiatric screening of flying personnel: Personality structure in objective tests—a study of 1,000 Air Force students in basic pilot training. *USAF Sch. Aviat. Med. proj. Rep.*, 1955, No. 9. (Project No. 21-0202-007) (b)

CATTELL, R. B. Second-order personality factors in the questionnaire realm. *J. consult. Psychol.*, 1956, **20**, 411–418.

CATTELL, R. B. *Personality and motivation structure and measurement.* New York: World Book, 1957. (a)

CATTELL, R. B. A universal index for psychological factors. *Psychologia*, 1957, **1**, 74–85. (b)

CATTELL, R. B. What is "objective" in "objective personality tests"? *J. counsel. Psychol.*, 1958, **5**, 285–289.

CATTELL, R. B., SAUNDERS, D. R., & STICE, G. *Handbook for the Sixteen Personality Factor Questionnaire.* Champaign, Ill.: Institute of Personality and Ability Testing, 1957.

CATTELL, R. B., & SCHEIER, I. H. Extension of meaning of objective test personality factors: Especially into anxiety, neuroticism, questionnaire, and physical factors. *J. gen. Psychol.*, 1959, **61**, 287–315.

CATTELL, R. B., & SCHEIER, I. H. *The meaning and measurement of neuroticism and anxiety*. New York: Ronald Press, 1961.

CATTELL, R. B., STICE, G. F., & KRISTY, N. F. A first approximation to nature-nurture ratios for eleven primary personality factors in objective tests. *J. abnorm. soc. Psychol.*, 1957, **54**, 143–159.

SCHEIER, I. H. What is an "objective" test? *Psychol. Rep.*, 1958, **4**, 147–157.

SCHEIER, I. H., & CATTELL, R. B. Confirmation of objective test factors and assessment of their relation to questionnaire factors: A factor analysis of 113 rating, questionnaire and objective test measurements of personality. *J. ment. Sci.*, 1958, **104**, 608–624.

(Received July 13, 1960)

Journal of Consulting Psychology
1961, Vol. 25, No. 4, 353–359

LEVEL OF ASPIRATION IN HYPERTENSIVE CARDIAC PATIENTS COMPARED WITH NONHYPERTENSIVE CARDIAC PATIENTS WITH ARTERIOSCLEROTIC HEART DISEASE

DESMOND D. O'CONNELL
Veterans Administration Hospital, Wood, Wisconsin
AND RICHARD M. LUNDY
University of Wisconsin

The present study [1] is concerned with the application of the "level of aspiration" technique to the study of persons with cardiac disease. Since 1930, when Hoppe first described the level of aspiration phenomenon, only six studies cited in *Psychological Abstracts* relate to physical or physiological conditions. Two studies—one reported by Berkeley (1952) and one reported by Gerard and Phillips (1953)—found a reliable relationship between adrenal activity and level of aspiration scores. Little and Cohen (1951) found that asthmatic children showed significantly higher levels of aspiration than nonasthmatic. Hecht (1952) was able to differentiate ulcer patients from colitis patients on level of aspiration scores, the ulcer patients getting significantly higher "D scores." Scodell (1953) also found differences between peptic ulcer patients and a neurotic (nonulcer) control group on a level of aspiration test as well as on several other measures; the ulcer group had lower level of aspiration scores than the neurotic group. Raifman (1957) found that a group of ulcer patients showed significantly higher levels of aspiration than a group of normals and a group of neurotics. The present study will attempt to determine whether persons with hypertensive heart disease can be differentiated from persons with nonhypertensive arteriosclerotic heart disease on a level of aspiration task.

Perhaps the best known effort to describe distinctive personality characteristics of persons with cardiac disease is that of Dunbar (1948) who described personality profiles for several types of patient. Alexander (1950) considers that probably the most valid of Dunbar's profiles is that of the "coronary" [2] patient. Yet an objective study of 46 coronary patients performed more recently by Miles, Waldfogel, Barrabee, and Cobb (1954), using psychiatric interview, social history, and psychological tests, disagrees with or fails to confirm most of the specific characteristics attributed to coronary patients by Dunbar. The study did show, however, more strenuous work histories, with more physical and psychological stress and strain than the group of normals; in this respect the study tends to confirm that part of Dunbar's profile which pictures the coronary patient as a consistently striving person who works harder and longer

[1] The authors wish to express their appreciation to Jules Chase, Consultant in Cardiology at the Veterans Administration Hospital, Wood, Wisconsin, for a suggestion made in the original design of the study and for his critical reading of the manuscript from the medical viewpoint.

[2] It will be noted that the word "coronary" in the introductory section has been used in quotation marks. This is because the term needs clarification; "coronary" and "hypertensive" are not mutually exclusive types of heart disease. Dunbar noted, for example, that 27% of the patients with coronary occlusion in her study also had hypertension (Dunbar, 1948, p. 251). Many hypertensives may eventually have coronary occlusion or coronary insufficiency. Miles, Waldfogel, Barrabee, and Cobb (1954) in their study of coronaries, limited the study to cases in which there had actually been a coronary attack—occlusion of the coronary artery—with an absence of hypertension.

than the average adult (Dunbar, 1948, p. 307). Alexander (1950, p. 72) agrees with Dunbar on this characteristic.

The person with hypertensive cardiac disease is characterized by Dunbar as having in common with coronary patients a constant striving to subdue or surpass competitors, but as being different from coronary patients in that hypertensives have a greater fear of criticism, a greater fear of responsibility, a greater fear of falling short. They are more likely to choose occupations below their ability and are usually less successful than coronaries (Dunbar, 1948, p. 264).

If hypertension is preceded by a history of being unsuccessful, fearful of criticism, or of responsibility, and of falling short, as suggested by Dunbar, it may produce a different pattern of response on a level of aspiration test than would be produced by a more successful, less fearful history. There is some support for this hypothesis in the literature. Sears (1940), for example, has shown that chronically unsuccessful children, as judged by school achievement, show a different pattern of response from successful children on a level of aspiration test. The pattern was not necessarily one of lowered aspirations, however, as might be suggested by Dunbar's assertion that the hypertensive is more likely to choose occupations below his ability. In Sears' study the unsuccessful group produced a bimodal distribution, showing either an unrealistically higher level of aspiration or an unnecessarily low level of aspiration. The successful individual typically sets his goal near but slightly above his last previous performance. Kurt Lewin (Lewin, Dembo, Festinger, & Sears, 1944) in his review of the literature and again in 1948 affirms these as characteristic reactions of the successful and the unsuccessful individual in level of aspiration situations.

In a recent theoretical article concerned with level of aspiration and risk taking behavior, Atkinson (1957) has drawn a distinction between those responses, or persons, whose motivation is to achieve success, and those whose motivation is to avoid failure. Those who seek to achieve success typically chose a task near their achievement level, i.e., where the uncertainty of success or failure is greatest. Those who seek to avoid failure typically set their goals either very high or very low. Thus in a level of aspiration study we might be concerned with two kinds of responses: the achievement oriented and the failure oriented. These types of responses seem comparable to Scodel's "typical" and "atypical" responses. Similarly, Scodel (1953) divides the atypical responses into atypical high and atypical low responses.

If Dunbar is correct in her assertions about persons with coronary heart disease and those with hypertensive heart disease, the two groups should be similar when achievement oriented behavior is considered. The responses of the two groups in their failure oriented behavior, however, should differ, according to her view that the hypertensives are more fearful of failure. She also states that the direction of this failure oriented reaction is toward choosing atypically low levels of aspiration—in Atkinson and Scodel's terms.

Although the writings of Dunbar have given impetus to this general investigation, the findings of Miles et al. (1954), have cast considerable doubt about her specific hypotheses. The approach of the present investigation, however, is not to test Dunbar's specific hypotheses. Rather, we feel that the status of the theoretical positions is such that only an empirical approach is presently warranted. Thus, the questions asked in the present study are: do hypertensive heart cases differ from nonhypertensive, arteriosclerotic heart cases in (*a*) the number of achievement oriented, aspiration responses, and (*b*) the extent and direction of the failure oriented responses?

Method

Subjects

All subjects were adult male veterans between the ages of 32 and 65, hospitalized at Veterans Administration Hospital, Wood, Wisconsin. Patients in this hospital are from Wisconsin, Michigan, and Illinois; the largest percentage of patients are from the Milwaukee metropolitan area.

All new admissions to the hospital were reviewed and a list made of all those showing an admission diagnosis of heart disease or suspected heart disease. After the medical diagnostic procedures had been completed, the cases were discussed with the ward physician. Subjects with an established diagnosis of

"Hypertensive Heart Disease" were assigned to one group; subjects with an established diagnosis of "Arteriosclerotic Heart Disease" and an absence of hypertension were assigned to the second group. Subjects with other types of heart disease were eliminated from further consideration as not meeting the criteria of the defined groups. Among those meeting the definitional or categorical requirements no eliminations were made except subjects who were mentally or physically incapable of cooperating in the study (for example, subjects who have had a cardiovascular accident resulting in hemiplegia, or aphasia, or deterioration of mental functioning).

Medically there is always the possibility that a subject in the nonhypertensive arteriosclerotic group has at some time in the past been hypertensive. However, from the standpoint of research design, such a misclassification would not invalidate a positive finding. Thus, if there is a psychological variable associated with the presence of hypertension, and the present study indicates that the nonhypertensive group and the hypertensive group differ significantly on this variable, the significance level would be a minimal value; transfer of the misclassified case to the hypertensive group in which it belongs would only serve to increase the significance level.

Associated diseases found to be present coincident with the primary diagnoses of hypertensive or arteriosclerotic heart disease were of interest, but were not used in selecting the subjects because to do so would change the basis of selection and disturb the research design of the study.

Each subject was given 11 15-second trials on the Placing test of the Minnesota Rate of Manipulation Test. He was asked to set a goal prior to each trial. The test was presented to him as a test of his coordination.

This task was selected after consideration of a number of factors. The literature on the use of the level of aspiration technique reveals that a wide variety of tasks, both verbal and motor, have been used; tasks used, for example, include card sorting, cancellation, addition, pegboards, bowling games, target tests, and the Rotter Aspiration Board. A decision was made in favor of the Minnesota Placing test for the following principal reasons: (*a*) Hecht (1952) found significant differences between peptic ulcer and ulcerative colitis patients using the Purdue Pegboard; the Minnesota Placing test is a somewhat similar performance-type test of coordination but requires less finger dexterity and less precise placement. (*b*) Bowling games and other "game"-type tasks in level of aspiration studies have been criticized by Barnett, Handelsman, Stewart, and Super (1952), Stubbins (1950), and others as having too much of a "play" atmosphere and not being closely related to life situations. This has led to the practice in recent level of aspiration studies of introducing penalties or other extraneous motivation. If this criticism has any validity, it might be that a game-type task would be even less meaningful and challenging to the older individuals with which the present study deals. On the other hand, it was our feeling that the Minnesota Placing test would prove to be a meaningful task when presented as "a test of coordination" to a hospitalized cardiac patient who has just come through a cautiously graded progression in amount of physical activity permitted and who is very much aware of and concerned about his physical limitations.

Test Instructions

Standardization of instructions was considered especially important in this study because it is known from previous studies of level of aspiration technique, such as those of Irwin (1942), of Walder (1951), and of Saji (1951) that the form of the stimulus question affects the results. Thus, "What score do you expect to get?" "What do you hope to get?" or "What will you try to get?" may bring different responses. Using Frank's (1941) definition of level of aspiration as, "the level of future performance in a familiar situation which an individual *explicitly* undertakes to reach," we have used as our stimulus question: "How many blocks will you try to get in on the next trial?" After the initial trial, the stimulus question was shortened to: "How many will you try for this time?"

Age

Age has not been shown to be a significant factor in level of aspiration studies; studies of this factor which have been done, such as those by Adams (1939), Walter (1948; Walter & Marzolf, 1951), and Reissman (1953), have tended to show that level of aspiration behavior is not significantly related to age. However, some of the studies did not examine age as a continuous variable but merely compared "old" groups with "young" groups; other studies which did examine age as a continuous variable have usually considered only a very limited range of ages. Because there has been insufficient research on age as a variable in level of aspiration behavior, and because age is known medically to be related to arteriosclerotic heart disease, it was our feeling that it should be controlled in the present study. The age of each subject was therefore recorded for comparison (see Results section).

Stage of Recovery

Although the effect of stage of recovery on level of aspiration behavior is unknown and uninvestigated, it seemed desirable to control this factor. The experimental task was light sedentary activity involving only 15 seconds of activity at a time and an over-all time of only about 15 minutes; yet it required rapid movement of the arm, and might be either a real or a perceived threat in the early stages of recovery from cardiac disease. Thus no patients were scheduled until they reached the "ambulatory" stage of recovery, and this was approximately the same for all subjects.

Education

The average education for the Hypertensive group is 8.7 with a standard deviation of 2.0, and a range of 5–13 years of formal schooling completed. This

TABLE 1

DIFFERENCE BETWEEN THE MEANS OF THE HYPERTENSIVE AND ARTERIOSCLEROTIC GROUPS USING THE "A" SCORES

Group	*N*	*M*	*SD*	SE_D	*t*
Hypertensive	24	5.75	2.57	0.72	1.83*
Arteriosclerotic	23	4.43	2.34		

* $p < .10$.

compares with an average of 8.8 and a standard deviation of 2.6 for the Arteriosclerotic group, with a range of 4–14 years completed.

Method of Scoring the Aspiration Responses

Although many systems of scoring have been used in various studies in the literature, the most widely used is the D score or the average discrepancy between the subject's performances and his aspirations. The D score, however, fails to take account of what would appear to be an important psychological variable: the effect of success or failure upon the subject's immediately subsequent aspiration. Thus, an average D score of +1 for one person may represent a very consistent aspiration to do one better than the last performance no matter whether it was a success or failure experience. For another person a +1 score could represent an average of quite different reactions following success or failure.

Guided by the Atkinson (1957) formulation that the achievement oriented individual sets his goal at the point of greatest uncertainty of success or failure, and assuming that this point is slightly above his last previous performance, we have set up the following operational definitions for the present study: achievement oriented aspiration—one point above the last previous performance; failure oriented high aspiration—two or more points above the last previous performance; failure oriented low aspiration—a goal at or below the last previous performance. The performances of the subject are operationally classified as follows: Success—reaching or surpassing one's goal; Partial Failure—a performance one point below the goal set; Failure—a performance two or more points below the goal set.

From the responses, two scores are derived: (*a*) An achievement oriented score designated as the "A score"; this is the total number of achievement-oriented responses. (*b*) A failure oriented score designated as the "F score"; this is the algebraic sum of the failure oriented responses weighted in the following manner: a high aspiration response following a Success was given a weight of 1; a high aspiration response following a Partial Failure was given a weight of 2; a high aspiration response following a Failure was given a weight of 3. Similarly, a low aspiration following a Failure was assigned a weight of −1; a low aspiration response following a Partial Failure was assigned a weight of −2, and a low aspiration response following a Success was assigned a weight of −3.

RESULTS

Table 1 presents results of a *t* test comparing the achievement oriented scores of the Hypertensive group with those of the nonhypertensive Arteriosclerotic group. The difference between the means is not statistically significant, although the direction of the difference shows a higher mean for the Hypertensive group.

Table 2 presents the results of a *t* test comparing failure oriented scores of the Hypertensive group with those of the nonhypertensive Arteriosclerotic group. The difference between the means was significant beyond the .05 level. A mean of 0.38 was found for the Hypertensive group, indicating that this group gave high aspiration responses about as often as low aspiration responses. A mean of −7.61 was found for the nonhypertensive Arteriosclerotic group, indicating that this group predominantly gave low aspiration responses.

Age has not generally been shown to be significantly related to level of aspiration. However, since it is known medically that it is related to arteriosclerosis, the question was raised as to the possibility that age had affected the results. However, analysis revealed identical means for the groups. The average age of the Arteriosclerotic group was 54.17, with a standard deviation of 8.43; the average age for the Hypertensive group was 54.13, with a standard deviation of 8.94. The ages ranged from 32 to 63 in the Arteriosclerotic group and from 34 to 65 in the Hypertensive group. The age factor, then, is not significant in the present study.

TABLE 2

DIFFERENCE BETWEEN THE MEANS OF THE HYPERTENSIVE AND ARTERIOSCLEROTIC GROUPS USING THE *F* SCORES

Group	*N*	*M*	*SD*	SE_D	*t*
Hypertensive	24	+0.38	12.45	3.54	2.26**
Arteriosclerotic	23	−7.61	11.79		

** $p < .05$.

Table 3 presents a comparison of the performance scores of the two groups. The mean of the Arteriosclerotic group was 13.91 with a standard deviation of 2.61; the mean of the Hypertensive group was 12.88 with a Standard deviation of 2.04. The difference of 1.03 in the means was not statistically significant.

Table 4 presents a comparison of the degree of improvement of the two groups as measured by the difference between the score on the first trial and the highest score attained on any of the subsequent trials. For the Arteriosclerotic group, the mean improvement was 3.48 with a standard deviation of 1.24. The mean for the Hypertensive group was 3.29 with a standard deviation of 1.28. The slight difference in the mean improvement (.19) was not statistically significant.

Discussion

The intent of this study was to determine whether a group of hypertensive cardiac patients would differ from a group of non-hypertensive arteriosclerotic cardiac patients on a level of aspiration test. The results show that when the level of aspiration is measured in terms of failure oriented responses, the two groups differ significantly with the hypertensives showing a higher level of aspiration response. In Atkinson's (1957) terms there are two ways of responding to fear of failure: choosing a goal in excess of reasonable expectation, or choosing a goal which can be easily attained. Thus, to defend against anxiety, the failure oriented individual either chooses a goal so high that failure to attain the goal is not destructive, or chooses a goal which is low enough so that he can be sure of achieving it. In terms of fear of failure, our data, therefore, suggest that in avoiding failure, the arteriosclerotic is more likely than the hypertensive to utilize the defense of choosing an easily attainable goal.

TABLE 3

Performance Scores of the Hypertensive and Arteriosclerotic Groups

Group	*N*	*M*	*SD*	SE_D	*t*
Hypertensive	24	12.88	2.04	0.65	1.57
Arteriosclerotic	23	13.91	2.61		

TABLE 4

Degree of Improvement of the Hypertensive and Arteriosclerotic Groups

Group	*N*	*M*	*SD*	SE_D	*t*
Hypertensive	24	3.29	1.28	0.38	0.50
Arteriosclerotic	23	3.48	1.24		

Analysis of the *frequency* of the achievement oriented responses fails to show a statistically significant difference between the groups. However, the significant finding is that when the failure oriented responses of both groups are analyzed, the hypertensive was more likely than the arteriosclerotic to choose a response which was so high as to assure failure. The difference between the arteriosclerotic and the hypertensive lies in the results of their different approaches to failure. The arteriosclerotic avoids the personal feeling of failure by "arranging" for objective success, i.e., by setting his aspiration low. The hypertensive, by contrast, arranges for objective failure. Thus no matter what the reason, the hypertensive less often reaches his goal—that is, he objectively fails.

The results of this study lead us to speculate about a possible relationship between hypertension and continual self-chosen failure. Reiser, Brust, and Ferris (1951) have discussed the role of "life stress" in the development of hypertension, assuming that the patient's reaction to the stress is elevated blood pressure. But just what, in the life stress of the hypertensive, would lead to the blood pressure elevation? Further, what in the life stress distinguishes the hypertensive from the arteriosclerotic, for the psychosomatic literature (Miles et al., 1954) includes studies which find that all heart patients live under more psychological stress than normals.

Our data would support the hypothesis that continued failure, resulting from too high goals, rather than a general life stress, can lead to hypertensive reactions. In at-

tempting to explain this relationship, we hypothesize that continued failure is in some way associated with a physiological withdrawal of blood from the arteriolar bed. The increased blood pressure of the hypertensive follows from the peripheral resistance in the arteriolar bed. Thus the hypertensive reaction is not the direct result of general psychological stress, but rather, is the direct result of the hypertensive avoidance reaction to the specific stress of continued failure. Support for the notion that the hypertensive reaction, i.e., constriction of peripheral vessels, is not due to general psychological stress itself comes from a well-controlled study by Baker and Taylor (1954) which found that skin temperature rises, i.e., arterioles dilate, under general psychological stress.

The present results also force some modification of Dunbar's characterization of the hypertensive. Hypertensives may fear falling short, as she suggests, but they react to this fear, not by choosing goals too low, as she suggests, but by choosing goals so high that they cannot possibly be achieved. Thus they insure failure.

Because this is the first study, to our knowledge, to apply level of aspiration techniques to the study of psychological variables in heart disease, the results obtained suggest further study of cardiac patients with this technique.

Summary

Twenty-four hypertensive cardiac patients and 23 nonhypertensive arteriosclerotic cardiac patients hospitalized at a Veterans Administration Hospital were administered a level of aspiration task based on the Minnesota Rate of Manipulation Test. Achievement oriented scores and failure oriented scores were derived from the aspiration responses. An achievement oriented response was operationally defined as an aspiration one point above the last previous performance; a failure oriented response was defined as either "high" (two or more points above last previous performance), or "low" (at or below the last previous performance). No significant difference between the groups was found in the frequency of achievement oriented scores. However, when failure oriented responses were broken down into high and low response patterns, the hypertensive group gave significantly ($p < .05$) less low aspiration responses than the nonhypertensive arteriosclerotic group. Level of performance and degree of improvement were not significantly different for the two groups. The hypertensive group "arranged" for repeated failure by consistently setting excessively high goals. It is hypothesized that the withdrawal of blood from the arteriolar bed, resulting in increased blood pressure, is an avoidance reaction to the repeated and continual failure experiences which the failure oriented hypertensive arranges for himself in a neurotic fashion.

References

Adams, D. K. Age, race, and responsiveness of levels of aspiration to success and failure. *Psychol. Bull.*, 1939, **36**, 573. (Abstract)

Alexander, F. *Psychosomatic medicine.* New York: Norton, 1950.

Atkinson, J. W. Motivational determinants of risk-taking behavior. *Psychol. Rev.*, 1957, **64**, 359–372.

Baker, L. M., & Taylor, W. M. The relationship under stress between changes in skin temperature, electrical skin resistance, and pulse rate. *J. exp. Psychol.*, 1954, **48**, 361–366.

Barnett, G. J., Handelsman, I., Stewart, L. H., & Super, D. E. The occupational level scale as a measure of drive. *Psychol. Monogr.*, 1952, **66**(10, Whole No. 342).

Berkeley, A. W. Level of aspiration in relation to adrenal cortical activity and the concept of stress. *J. comp. physiol. Psychol.*, 1952, **45**, 443–449.

Dunbar, F. *Psychosomatic diagnosis.* New York: Harper, 1948.

Frank, J. D. Recent studies of the level of aspiration. *Psychol. Bull.*, 1941, **38**, 218–225.

Gerard, D. L., & Phillips, L. Relation of social attainment to psychological and adrenocortical reactions to stress. *AMA Arch. Neurol. Psychiat.*, 1953, **69**, 350–354.

Hecht, I. The difference in goal striving behavior between peptic ulcer and ulcerative colitis patients as evaluated by psychological techniques. *J. clin. Psychol.*, 1952, **8**, 262–265.

Irwin, F. W., & Mintzer, M. G. Effect of differences in instructions and motivation upon measures of the level of aspiration. *Amer. J. Psychol.*, 1942, **55**, 400–406.

Lewin, K. *Resolving social conflicts.* New York: Harper, 1948.

Lewin, K., Dembo, Tamara, Festinger, L., & Sears, Pauline. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. 1. New York: Ronald, 1944. Pp. 333–378.

LITTLE, SUE W., & COHEN, L. D. Goal setting behavior of asthmatic children and of their mothers for them. *J. Pers.*, 1951, **19**, 376–389.

MILES, H. H. W., WALDFOGEL, S., BARRABEE, EDNA L., & COBB, S. Psychosomatic study of 46 young men with coronary artery disease. *Psychosom. Med.*, 1954, **16**, 455–477.

RAIFMAN, I. Level of aspiration in a group of peptic ulcer patients. *J. consult. Psychol.*, 1957, **21**, 229–231.

REISER, M. F., BRUST, A. A., & FERRIS, E. B., JR. Life situations, emotions and the course of patients with arterial hypertension. *Psychosom. Med.*, 1951, **13**, 133.

REISSMAN, L. Levels of aspiration and social class. *Amer. sociol. Rev.*, 1953, **18**, 233–242.

SAJI, M. The degree of reality in level of aspiration. *Jap. J. Psychol.*, 1951, **21**(3–4), 56–69.

SCODELL, A. Passivity in a class of peptic ulcer patients. *Psychol. Monogr.*, 1953, **67**(10, Whole No. 360).

SEARS, P. S. Levels of aspiration in academically successful and unsuccessful children. *J. abnorm. soc. Psychol.*, 1940, **35**, 498–536.

STUBBINS, J. The relationship between level of vocational aspiration and certain personal data: A study of some traits and influences bearing on the prestige level of vocational choice. *Genet. psychol. Monogr.*, 1950, **41**, 327–408.

WALDER, L. O. The effects of instructions, order, and sex-ethnic group in level of aspiration situations. Unpublished masters dissertation, University of Hawaii, 1951.

WALTER, L. M. The relation of age and sex to levels of aspiration. Unpublished masters dissertation, Illinois State Normal University, 1948.

WALTER, L. M., & MARZOLF, S. S. The relation of sex, age, and school achievement to levels of aspiration. *J. educ. Psychol.*, 1951, **42**, 285–292.

(Received July 20, 1960)

Journal of Consulting Psychology
1961, Vol. 25, No. 4, 360–361

OBJECTIVE ESTIMATES OF CLINICAL JUDGMENTS[1]

ROLFE LAFORGE
University of Illinois

One of the more promising avenues towards a theory of personality develops from the discovery of the kind of structure most appropriate for a model of clinical inference. The present paper reports an early and imperfect example of this approach. There are at least four major imperfections.

First, while one would prefer to abstract representative tasks of clinical inference, in this study only one kind of information and only one clinical construct were available. MMPI profiles of scores on the 12 *K*-corrected clinical scales were given to four psychologists experienced in the use of the MMPI. Each psychologist rated the "degree to which *repression* was relied upon by the patient as a defense against anxiety." The average of these four ratings for any patient is the value of the "clinically inferred" variable (*R*).[2]

Second, one would like to ask the psychologists who make the clinical inferences to describe also the cues, working hypotheses, and methods of verification which they use. Here, each rater was asked only to list as many "cues" of repression as could be found in MMPI profiles. The four lists were pooled, and duplications removed; 27 cues remained, with many dependencies among them.

Third, one would like to examine the relations between data and construct, and between construct and prediction, using the most general statistical models available. In this study only two models were applied. One model employed a linear prediction of *R* from its multiple regression on *Hy* and *Sc*. The other method sought cues which could be associated with different ordinal subregions within the total range of *R*. That is, the variable *R* was used to order the profiles within a sample; then the Mann-Whitney and the Wald-Wolfowitz (runs) tests were applied separately to each of the suggested cues to test for relationship to *R*. A 5% level of significance was used in all tests. In the first sample of 35 outpatients, eight cues were related to *R* by the Mann-Whitney. Four of these same cues were related to *R* according to the runs test, plus two additional cues which had a nonmonotonic relationship with *R*. (The four found by the Mann-Whitney alone are presumably the result of its greater power against monotonic alternatives to the null hypothesis.)

Fourth, and most serious of the imperfections, cross-validation of these results on the second sample of 83 outpatients became impossible as a result of loss of data during a 6-year lapse. The investigator redefined from memory 26 cues whose relation to *R* was then tested in the second sample by the runs test alone. Of 14 "pattern" cues referring to the relative height of two or more scales, 11 were found to relate to *R;* of 11 "absolute elevation" cues, 4 related to *R*. One "mixed" cue, referring to both elevation and pattern, also predicted *R*.

These 16 cues could be used as a checklist to enable the MMPI novice or researcher to make a quick estimate of *R* from an MMPI profile. To simplify the estimation process, a Guttman scale was constructed by eliminating cues approximately equal in relative frequency to others. The remaining seven cues scaled in the second sample with a reproducibility above .90; removal of one cue raised this to .96. A check on a third sample of 201 outpatients gave a reproducibility of .93. Be-

[1] This work was supported by United States Public Health Service Postdoctoral Fellowship M3634 (1952).

[2] For details, see: LaForge, Leary, Naboisek, Coffey, Freedman (1954, pp. 132, 135–136).

cause the cues are to some extent experimentally dependent, this figure indicates only the sufficiency of the estimate in use. The six-item Guttman-type scale [3] correlated as well with *R* (.799) as did the simple sum over all predictive items (.788). In comparison, the multiple correlation of *Hy* and *Sc* with *R*, based on the same sample of 83 cases, was .843. (The beta weights in this sample were, for *K*-corrected *T* scores: *Hy*, .0718; *Sc*, −.0460. Thus a quick estimate of *R* could be obtained by subtracting *Sc* from $1\frac{1}{2}$ times *Hy*.)

On the basis of the present evidence, there is little to choose between the two estimates. Of course the lack of cross-validation makes comparison particularly difficult. Consideration of those cues which predicted *R* as against those which did not significantly discriminate suggests that the emphasis on configuration or pattern in the teaching of clinical inference may be justified. On the other hand, the equal success of a linear combination of two scales points up the psychometrician's faith in the efficacy of negatively correlated ($r_{HySc} = -.379$) predictors in an apparently configural situation. And either objective measure showed correlations with *R* approximately as high as the interrater correlations.

[3] The Guttman-scaled cues are $L \geqslant 60$, $L \geqslant F$, $Hs + Hy \geqslant D + Sc$, $Hy \geqslant Sc$, $K + 10 \geqslant F$, and $Hy + 10 \geqslant D$, in order of increasing frequency of occurrence. The most predictive single cue was $Hs + Hy \geqslant D + Sc$.

REFERENCE

LaForge, R., Leary, T. F., Naboisek, H., Coffey, H. S., & Freedman, M. B. The interpersonal dimension of personality: II. An objective study of repression. *J. Pers.*, 1954, 23, 129–153.

(Received June 20, 1960)

Journal of Consulting Psychology
1961, Vol. 25, No. 4, 362–363

A NOTE ON "IMPULSE REPRESSION AND EMOTIONAL ADJUSTMENT"

H. J. EYSENCK
University of London

In a recent paper, Grater (1960) has tested the Freudian theory of impulse repression as a correlate of neuroticism, emerging with conclusions which apparently contradict that theory; the more neurotic subjects were, if anything, less "repressive" than were the nonneurotic ones. This result would appear to be in line more with Mowrer's (1953) view of neurosis, according to which

> the problem-solving activity which is usually referred to clinically as self-protectiveness or defensiveness . . . functions in the interest of the primary drives or id, rather than, as Freud posited, in the services of the socially derived forces of the superego (p. 145).

I have discussed the point at issue between orthodox Freudian writers and Mowrer elsewhere (Eysenck, 1957, p. 82f.), and have suggested there that the distinction which must be made in order to accommodate the known facts is one between *extraverted* neurotic behavior patterns (hysteria, psychopathy, hypochondriasis, etc.) and *introverted* neurotic behavior patterns (anxiety, reactive depression, obsessional-compulsive, dysthymic reactions). This personality dimension of extraversion-introversion is conceived of as being orthogonal to neuroticism, and I have further suggested that "impulse repression" and socialization generally are in part caused by constitutional factors closely linked with introversion. Dysthymic neurotics, according to this view, are "oversocialized," hysteric and psychopathic ones "undersocialized." This theory has been discussed in some detail in relation to the experimental evidence (Eysenck, 1957, 1960a, 1960c) and it may be concluded that it serves to reconcile a large amount of factual material.

When we turn to Grater's study we find that he has defined neuroticism in terms of three MMPI scales two of which are measures of extraverted neuroticism (*Hy, Hs*), while the most clear-cut introverted scale (*Pt*) was not used at all. According to the analysis given above, therefore, we would expect neurotics (as defined by Grater's MMPI scores) to be extraverted and less given to impulse repression than nonneurotics. His results, as far as they go, bear out this prediction, although in only one or two instances do his scores reach statistical significance.

The purpose of this note is not so much to reinterpret Grater's data as to draw attention to the absolute necessity, in work of this kind, to take into account the two-dimensional nature of the test-space in which the experimenter is working (Eysenck, 1960b). Much experimental work in this field is wasted because results are quite uninterpretable, it being impossible from the data given to sort out the dimensions involved; work with the Manifest Anxiety scale is a good example of this, the resulting score having loadings both on neuroticism and on introversion (Bendig, 1960; Eysenck, 1957). Much of the theoretical disputation regarding the nature of neuroticism is sidetracked by emphasizing either the extraverted or the introverted side (Miller & Dollard, 1950; Mowrer, 1953). The evidence for *at least* two factors in this field is now practically conclusive (Eysenck, 1960b) and it would seem desirable to recognize this fact in the design and interpretation of psychological experiments.

REFERENCES

BENDIG, A. W. Factor analysis of "anxiety" and "neuroticism" inventories. *J. consult. Psychol.*, 1960, **24**, 161–168.

EYSENCK, H. J. *The dynamics of anxiety and hysteria*. London: Routledge & Kegan Paul, 1957.

EYSENCK, H. J. *Experiments in personality*. London: Routledge & Kegan Paul, 1960. 2 vols. (a)

EYSENCK, H. J. *The structure of human personality*. (2nd ed.) London: Methuen, 1960. (b)

EYSENCK, H. J. Symposium: The development of moral values in children: VII. The contribution of learning theory. *Brit. J. educ. Psychol.*, 1960, **30**, 11–21. (c)

GRATER, H. A. Impulse repression and emotional adjustment. *J. consult. Psychol.*, 1960, **24**, 144–149.

MILLER, N. E., & DOLLARD, J. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.

MOWRER, O. H. *Psychotherapy: Theory and research*. New York: Ronald, 1953.

(Received July 1, 1960)

Journal of Consulting Psychology
1961, Vol. 25, No. 4, 364–365

SEX DIFFERENCES IN MENTAL HEALTH ANALYSIS SCORES OF ELEMENTARY PUPILS

JOSEPH C. BLEDSOE
University of Georgia

The manner in which the individual perceives himself has been regarded as one indicator of the degree of mental health he demonstrates or possesses. Among the currently available instruments for measuring self-perception of mental health status is the Mental Health Analysis prepared by the California Test Bureau. The test manual makes no mention of sex differences in norms for the Mental Health Analysis. Previous studies (Ausubel, Balthazar, Rosenthal, Blackman, Schpoont, & Welkowitz, 1955; Davidson, Sarason, Lighthall, Waite, & Sarnoff, 1958; Sarason, Davidson, Lighthall, & Waite, 1958) have indicated that girls may perceive themselves as significantly more accepted and intrinsically valued than boys. The present study provides more evidence of the possibility of sex differences in mental health status, as reflected by responses to the Mental Health Analysis.

The subjects were 96 girls and 101 boys enrolled in the fourth through the seventh grades in an elementary school of a southeastern city of approximately 35,000 population. The Elementary Form of the Mental Health Analysis was administered as a part of a 3-year in-service education program for teachers designed to promote better understanding of mental health principles and practices.

The Mental Health Analysis consists of 200 items to be answered by circling Yes or No. Five sorts of personality "liabilities" and five sorts of "assets" are investigated. Liabilities include behavioral immaturity, emotional instability, feelings of inadequacy, physical defects, and nervous manifestations. The assets include close personal relationships, interpersonal skills, social participation, satisfying work and recreation, and adequate outlook and goals. The reliability of the Elementary scale is reported as: assets, .90; liabilities, .89; total, .90. Reliability coefficients for component scores of the Elementary scale vary from .80 for "outlook and goals" to .85 for "physical defects." Content validity of the instrument is based upon the adequacy of item selection, the meaningfulness of the analysis of the mental health categories, and the cleverness in disguise of items. Studies of concurrent validity are reported in the test manual (California Test Bureau, 1959).

Means and standard deviations for the several components, categories, and total scores on the Mental Health Analysis for boys and girls were computed.[1] Differences in means were then tested for significance by the *t* test of significance. The .05 level of confidence was accepted as the criterion for rejection of the null hypotheses involved. Twelve of the 13 differences favored the girls, and 7 of these differences met the .05 level of significance criterion. Significant differences were found in the total score, the total assets, the total liabilities, and in close personal relationships, outlook and goals (assets subscales), and behavioral immaturity and feelings of inadequacy (liabilities subscales). A nonsignificant difference favoring the boys was found in the physical defects subscale of the liabilities category. Thus within the limitations of the study, it appears that elementary school girls tend to rate themselves significantly higher on

[1] A summary table of these statistics has been deposited with the American Documentation Institute. Order Document No. 6762 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

the Mental Health Analysis than do elementary age boys. Implications for differential norms may be suggested. More important may be implications for teacher understanding of and curriculum adjustment to the needs of boys.

REFERENCES

AUSUBEL, D. P., BALTHAZAR, E. E., ROSENTHAL, I., BLACKMAN, L. S., SCHPOONT, S. N., & WELKOWITZ, J. Perceived parent attitudes as determinants of children's ego structure. *Child Develpm.*, 1955, **25**, 173–183.

CALIFORNIA TEST BUREAU. *Manual for Mental Health Analysis.* (Rev. ed.) Los Angeles: CTB, 1959.

DAVIDSON, K. S., SARASON, S. B., LIGHTHALL, F. F., WAITE, R. R., & SARNOFF, I. Differences between mothers' and fathers' ratings of low anxious and high anxious children. *Child Develpm.*, 1958, **29**, 155–160.

SARASON, S. B., DAVIDSON, K., LIGHTHALL, F., & WAITE, R. Rorschach behavior and performance of high and low anxious children. *Child Develpm.*, 1958, **20**, 277–285.

(Early publication received December 12, 1960)

Journal of Consulting Psychology
1961, Vol. 25, No. 4, 366

BRIEF REPORTS

A COMPARISON OF ACCEPTORS AND RESISTORS OF DRUG TREATMENT AS AN ADJUNCT TO PSYCHOTHERAPY [1]

ALLEN RASKIN

Veterans Administration, Washington, D. C.

In a recent adjunct chemotherapy study with psychiatric outpatients, two of the major causes of patient attrition were refusal to take the study medication and excessive deviation from prescribed dosage levels. Can these patients be identified early in treatment and why are they reluctant to take the assigned medication? This study was undertaken as a preliminary effort to answer these questions by identifying variables which differentiate patients who remained in psychotherapy but resisted taking drugs from patients who accepted drug treatment as an adjunct to psychotherapy.

Data for the present study were collected in 22 Veterans Administration Mental Hygiene Clinics as part of a larger chemotherapy study. There were four psychotherapy-plus-drug groups and one psychotherapy-only group in the larger study. The four study drugs were chlorpromazine, meprobamate, phenobarbital, and placebo. The study was conducted double-blind. Patients were told that the drugs had helped a lot of people with similar troubles. Study patients were all males, under age 50, who were acceptable for individual psychotherapy. There were 142 patients (the Acceptors) who remained in psychotherapy at least 8 weeks and took their medication as prescribed. An additional 37 patients (the Resistors) also remained in psychotherapy for at least 8 weeks but either refused to take the assigned medication or took significantly less than the amount prescribed. As Acceptors and Resistors did not differ significantly by treatment group, all Acceptors were pooled into one group and all Resistors into another.

On the basis of data obtained from the patients and from their therapists, the Acceptors and Resistors were compared on 40 personality, socioeconomic, and attitudinal variables. Resistors did not report a greater number of adverse side effects at the end of the first week on medication. Eleven significant differences were found between these two groups as compared to two expected by chance for the number of tests made. Compared to the Acceptors, the Resistors were better educated, had a greater knowledge of psychiatry, had less favorable attitudes toward physicians, rated themselves more hostile on an adjective checklist, and admitted to greater direct expression of hostility. At the end of the initial interview, therapists rated Resistors as expressing a more negative attitude toward taking the assigned drugs, less likely to have their psychotherapy facilitated by the addition of a tranquilizer, more inwardly hostile, reporting more incidents of overt aggression, and less likable than Acceptors. After 8 weeks of treatment, therapists also rated the Resistors as more resistive to psychotherapy.

These findings indicate that by the end of the initial psychotherapy session, therapists were able to identify many Resistors by their negative attitude toward taking the assigned medication. The Resistors' reluctance to take the study drugs was not due to side effects of the drugs. Instead, the drugs apparently provided a convenient focal point for the hostile and aggressive impulses noted by the Resistors' therapists and admitted by the patients themselves.

(Received September 23, 1960)

[1] An extended report of this study may be obtained without charge from Allen Raskin (Neuropsychiatric Research Laboratory, Veterans Benefits Office; Munitions Building; Washington 25, D. C.) or for a fee from the American Documentation Institute. Order Document No. 6647 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
1961, Vol. 25, No. 4, 367

AN INTERPRETATION OF *m* [1]

JOHN M. REISMAN
Rochester Child Guidance Clinic, New York

Piotrowski (1957) hypothesized that the types of movement in *m* responses are always different from those of *M* and *FM*. His hypothesis was based on his interpretations of *m* as tendencies which are never acted out while tendencies indicated by *M* and *FM* might be manifested in overt behavior. These "tendencies" were said to be primarily indicated by three types of action: flexor, extensor, and blocked movements. Respectively, they were characterized by giving in to gravity, overcoming gravity, and tension; they were interpreted as tendencies toward compliance, self-assertion, and indecisiveness.

Using a sample of disturbed children whose mean age was 10, Reisman (1960) found results directly opposite to those predicted by the hypothesis. This study repeated the test of the hypothesis with adolescents to determine if, with increasing age, there is a tendency for a person's *m* to differ in type from his *M* and *FM*.

Only responses that were identical with or highly similar to Piotrowski's examples of extensor, flexor, and blocked movements were considered. With this restriction, from a pool of 80 Rorschachs obtained from adolescents referred to a child guidance clinic, only 22 were found which contained at least 1 *m* and 1 *M*. In this sample, there were 18 boys and 4 girls. Ages ranged from 13–15; IQ scores ranged from 80–120 ($\bar{X} = 102$). None of the subjects was considered psychotic. All but one of the subjects, who was referred for underachieving in school, had been referred for acts of delinquency.

In a test for reliability, there was 90% and 80% agreement between the experimenter and two judges in categorizing 96 responses as to type. Disagreements did not appreciably affect the results. Movement responses were extracted from their records, coded, and written verbatim in a random order. Three months elapsed between the compilation and the categorizing of responses.

Subjects produced a scorable total of 29 extensor, 19 flexor, and 2 blocked *M* responses; 30 extensor, 6 flexor, and 0 blocked *FM* responses; and 25 extensor, 2 flexor, and 1 blocked *m* responses. These results were similar to those of the previous study and to findings with a sample of "normal" adolescents. Of the 22 records, 18 had *m* and *M* or *FM* responses of the same type; 14 records had *m* and *M* responses of the same type. These results were, once again, in direct opposition to Piotrowski's hypothesis. Furthermore, if it had been predicted that an adolescent's *M* or *FM* is of the same type as his *m*, this hypothesis would have been supported ($\chi^2 = 7.68$; $p < .01$).

Piotrowski has stated that tendencies indicated by *m* are never acted out. In only one case, that of a boy whose *m* was extensor and who was referred for underachieving, was this expectation supported by the record of overt behavior. It would seem reasonable to conclude that, depending upon the adolescent's controls, the tendencies indicated by *m* may or may not be acted out. The results further suggest that *m* be interpreted as an awareness of impulses or feelings which the individual experiences difficulty in controlling and expressing. Piotrowski's types of movement seem to offer assistance in understanding what kinds of impulses or feelings cause the person concern.

[1] An extended report of this study may be obtained without charge from John Reisman (Rochester Child Guidance Clinic; 31 Gibbs Street; Rochester 4, New York) or for a fee from the American Documentation Institute. Order Document No. 6648 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

REFERENCES

PIOTROWSKI, Z. N. *Perceptanalysis*. New York: Macmillan, 1957.

REISMAN, J. M. Types of movement in children's Rorschachs. *J. proj. Tech.*, 1960, **24**, 46–48.

(Received October 7, 1960)

Journal of Consulting Psychology
1961, Vol. 25, No. 4, 368

THE RELATION OF THE FAMOUS SAYINGS TEST TO SELF- AND IDEAL-SELF-ADJUSTMENT [1]

BERNARD I. MURSTEIN

Interfaith Counseling Center, Portland, Oregon

Bass (1958, p. 479) cites Thurstone as stating "the best form of projective test is one which is quite unstructured for the subject but fairly well structured for the examiner." Using Thurstone's words as a model, Bass attempted to construct a new projective technique in which administration and scoring are completely objective. The subject is presented with a booklet of 100 proverbs which he answers by checking the Yes, ?, or No box for each proverb. From his answers four slightly correlated scales have been derived. These are Social Acquiescence, Fear of Failure, Conventional Mores, and Hostility. Social Acquiescence has received wide publicity because of the role it is reputed to play in most personality questionnaires. Little, however, has been done to study the meaning of Social Acquiescence apart from its role as an interferer in the study of other variables. Accordingly, to study the personality correlates of the Famous Sayings Test (FST), scores on this test were compared to a personality questionnaire adaptation of the Butler-Haigh SIO *Q* sort instrument.

The hypothesis for this study was that there is no significant correlation between the scales of the FST and the following scales of the SIO *Q* sort: (*a*) Self-Adjustment, (*b*) Ideal-Self-Adjustment, and (*c*) Self–Ideal-Self-Adjustment Discrepancy.

Forty-five subjects (24M, 21F) from the introductory psychology class at the University of Portland were told that the experimenter wished to investigate the merits of a new type of questionnaire about how people describe themselves as well as to construct a test of attitudes towards famous sayings.

Of 15 correlations bearing on the hypothesis, only the correlation between Social Acquiescence and Self–Ideal-Self-Adjustment Discrepancy proved to be significant at the .01 level ($r = .45$). The results lend support to the interpretation of Social Acquiescence as more than an interfering set. Social Acquiescence may be a superficial phenotype reflecting considerable underlying anxiety which manifests itself by an inordinate need to win acceptance by others through excessive conformity.

As to the other scales—Conventional Mores, Hostility, and Fear of Failure—none showed any relation to Self- or Ideal-Self-Adjustment. Considering earlier failures to find meaningful correlates for these scales, it must be considered doubtful that they will eventually find a niche for themselves in trait measurement. The reason for this pessimism is that, in addition to rather low reliabilities, these scales make the further, probably unwarranted, assumption that saying "yes" to a hostile proverb indicates that one is hostile. It must be remembered that proverbs are not necessarily simple, bland statements that have no consensual validation about their general truth. Accordingly, it is entirely possible that many of the proverbs may be perceived as being true by the majority of persons. Hence, there can be little differential value as to these scales unless scaled values can be established as to the normative agreement on the verification of a proverb. To a lesser degree the same criticism is applicable to Social Acquiescence. The fact that it contains 56 items compared to 30 for each of the others and is not limited to a specific content as is true of the others, probably accounted for its greater validity.

[1] An extended report of this study may be obtained without charge from Bernard Murstein (Interfaith Counseling Center; 729 Southwest Alder; Portland, Oregon) or for a fee from the American Documentation Institute. Order Document No. 6759 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

REFERENCE

BASS, B. M. Famous Sayings Test: General manual. *Psychol. Rep.*, 1958, **4**, 479–497.

(Received October 17, 1960)

Journal of Consulting Psychology
1961, Vol. 25, No. 4, 369

THE INTELLIGENCE OF BOYS WITH MUSCULAR DYSTROPHY[1]

DON KEITH WORDEN
Western Reserve University

Since pseudohypertrophic muscular dystrophy (PMD) was first described there has been controversy concerning the intelligence of the children who have it. Several studies done recently claim that these children do not suffer from mental deterioration, and that they are of average intelligence. Academic retardation has been mentioned as a frequent problem, in some of these studies.

This study was designed to clarify some of these points in a more rigorous manner than has been done to date. A group of boys with PMD was given intelligence and achievement tests.

Several further questions were raised in this investigation: (*a*) Would the measured depression in intelligence be partly a function of the child's socioeconomic background? (*b*) Does any child with a chronic disease show depression in intelligence? (*c*) Would measured depression in intelligence be due partly to the factor of the child's having a physically handicapping (as distinct from merely chronic) illness? Appropriate control groups were selected to investigate these questions for: (*a*) siblings of the PMD group, (*b*) diabetics, and (*c*) amyotonia congenita.

Form M of the 1937 Stanford-Binet scale was used as the measure of intelligence. The measures of academic achievement were the Gilmore Oral Reading Test and the Metropolitan Achievement Test in Arithmetic Fundamentals. The educational quotient (EA/CA × 100) and the accomplishment quotient (EA/MA × 100) were figured.

The muscular dystrophy group consisted of 38 school age boys. It was possible to test 27 of these children's siblings on the Binet.

The amyotonia congenita group consisted of 16 children. The diabetic and the diabetic-sibling group consisted of 36 children each.

The PMD group had a mean IQ of 83. The range was from 46 to 134, but skewed radically to the left.

Twenty-five of these 38 boys were given the achievement test in arithmetic. The mean educational quotient (EQ) of this group was 84. The accomplishment quotient (AQ) was 96. The reading test was administered to 24 of these 38 boys. The mean educational quotient for reading was 87 and the mean accomplishment quotient was 101. The IQ did not differ significantly from the EQ in either reading or arithmetic. These three measures, however, differed significantly from the AQ in reading and the AQ in arithmetic. The two AQs did not differ significantly from one another.

The siblings of the PMD group had a mean IQ of 110. The children with diabetes mellitus had a mean IQ of 107. The diabetic siblings had a mean IQ of 109. The patients with amyotonia congenita had a mean IQ of 118. The analysis indicated that the PMD group differed significantly from all control groups.

In summary, boys with pseudohypertrophic muscular dystrophy functioned on an intelligence test significantly below average and below several control groups. Their IQs and EQs are both below average and do not differ from one another. However their AQs indicate that they are doing all one could expect for children of their mental ages.

Thus, within the confines of this study, the intellectual deficit seems specifically associated with the factor of having PMD.

(Received October 31, 1960)

[1] An extended report of this study may be obtained without charge from Don K. Worden (Southeastern Ohio Guidance Center; Athens, Ohio) or for a fee from the American Documentation Institute. Order Document No. 6760 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
1961, Vol. 25, No. 4, 370

SYMBOLIC INTERPRETATION OF RORSCHACH CONTENT[1]

JOSEPH F. RYCHLAK AND DONALD E. GUINOUARD
St. Louis University *Montana State College*

This research continues the examination of 12 Rorschach contents studied earlier (Rychlak, 1959): Bat, Bear, Boots, Clouds, Fire, Fur, Hair, Island, Mask, Mountains, Rocks, and Smoke. It was decided that a culling of Rorschach protocols for these contents, and then a comparison made along personality dimensions for subjects perceiving certain of these contents might throw preliminary light on the little studied area of content symbolism.

Subjects were 80 girls and 86 boys in the 11–14 year age range. Testing was done in groups, with the Harrower (1959) Group Rorschach slides and the High School Personality Questionnaire (Cattell, Beloff, & Coan, 1958) used as inkblot and personality measures, respectively. All testing was completed within one week. Scoring percentage of agreement between judges for identification of the contents was 95; subjects' one week test-retest reliabilities—based on binomial expansion—reached the .01 level for half of the contents.

Frequently reported contents for all subjects ($N = 166$) included Bat, Bear, and Fire; whereas Fur, Hair, and Island were infrequently noted. Girls reported Clouds, Hair, and Smoke content significantly more often than boys ($p < .05$ or less). Boys were more likely than girls to report Mask content ($p < .01$).

The (chi square) .01 level personality findings for the total sample were: Subjects who see Boots tend to be talkative and cheerful. Children who report Bat content are less tense and excitable than their peers. The reporting of Fur is related to a form of emotional instability, suggesting overactivity and frustration. The .05 level findings for total sample were: Fire content is reported more frequently by subjects who are tense and excitable. Children who report Rocks content tend to be dominant, independent, and outgoing. Island content is suggestive of a tough, self-sufficient personality.

Considered individually by sex, the following .05 level findings were noted: Boys who see Clouds in the inkblots are phlegmatic, deliberate, and self-effacing; girls who see Clouds are tough, realistic, and group-conforming. Girls reporting Hair are self-concerned and individualistic. The contents Bear, Smoke, Mask, and Mountains did not discriminate between subjects' personalities.

To propose to study potential symbols nomothetically implies the existence of a kind of "universal" symbolism, something repugnant to the common sense of many clinicians. Yet, in broad terms, such a symbolism may be possible through the intervention of cultural factors. Reference to a collective unconscious by some theorists would be interpreted from this viewpoint to mean that there are variables in any culture which—though learned by all or by many—are unverbalized. They are the shadow side of cultural manifestation. The clinician, through studies of this sort and the assumption of the proper set as a result of his study, may identify these variables in his client's projective responses, dreams, etc. In this connection, interesting parallels were noted between the present findings and the forced associations of the earlier study (Rychlak, 1959).

[1] An extended report of this study may be obtained without charge from Joseph F. Rychlak (Department of Psychology, St. Louis University; St. Louis 3, Missouri) or for a fee from the American Documentation Institute. Order Document No. 6761 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

REFERENCES

CATTELL, R. B., BELOFF, H., & COAN, R. W. *High School Personality Questionnaire*. Champaign, Ill.: Institute for Personality & Ability Testing, 1958.

HARROWER, MOLLY R. *Group Rorschach slides*. New York: Psychological Corporation, 1959.

RYCHLAK, J. F. Forced associations, symbolism, and Rorschach constructs. *J. consult. Psychol.*, 1959, 23, 455–460.

(Received November 23, 1960)

PSYCHOLOGICAL TEST SPECIALISTS

Exclusive publishers and distributors of selected psychological tests

Memory-For-Designs Test (MFD)

Simple drawing test of perceptual-motor coordination, serving as extremely sensitive detector of brain injury of many types. $8.50 for set of materials

Group Personality Projective Test (GPPT)

Multiple-choice questions about ambiguous stick-figures make possible rapid and effective measurement of significant, factor-analytically isolated aspects of personality. $13.50 for set of materials

Id-Ego-Superego Test (IES)

Subtests give independent estimates of personality components from a modified psychoanalytic point of view. $23.50 for set of materials

Kahn Test of Symbol Arrangement (KTSA)

Uses symbolic plastic objects which are arranged and sorted; validly discriminates among normals, neurotics, brain-injured, and psychotics, and indicates accompanying personality dynamics. $25.00 for set of materials

Full-Range Picture Vocabulary Test (FRPV)

Brief test of intelligence in two forms using the same 16 plates. Words are defined by choosing pictures. Valuable for intelligence screening, and with the physically handicapped. $7.50 for set of materials

Geist Picture Interest Inventory (GPII)

Provides for more valid expression of interests by choices among alternative pictured activities. Both qualitative and quantitative interpretations are possible. $6.00 for set of materials

Proverbs Test (PT)

Group and individual forms determine similarity of thought processes to those of various normal and clinical groups. Highly sensitive indicator of psychotic processes. $4.50 for kit

Vocational Apperception Test: Advanced Form (VAT:ADV)

Separate series of plates showing male and female college-related occupational activities, accompanied by explanation of use and interpretation as a projective test. $7.25 for set of materials

Famous Sayings (FS)

Agreement with proverbs, aphorisms, and folk sayings is analyzed to determine personality structure. Particularly useful in personnel selection. $4.50 for kit

It Scale For Children (ITSC)

An experimental picture scale for measuring sex role preference in children. $15.00 for set of materials

Kahn Intelligence Tests: Experimental Form (KIT:EXP)

Relatively culture-free intelligence tests requiring no reading or writing. Supplementary scales for reassessment of intelligence of blind, ability in concept formation, recall, and motor coordination. One scale can be administered by sign language. $26.00 for set of materials

PSYCHOLOGICAL TEST SPECIALISTS

Box 1441 Missoula, Montana

WHAT

DISTINGUISHES

DEVEREUX

in its service to children who need remedial education? It furnishes —

1. Group living and learning experience with others of a similar aptitude and level of development.

2. The functioning of a multidisciplinary team with long experience in evaluating potential and in structuring programs in a residential setting unique in its wide range of homogeneous groupings.

3. A philosophy of optimum blending of traditional methods with the best of the new from the frontiers of research.

4. Established programs of diagnosis, treatment, research, and training soundly based on the wide spectrum of a multidisciplinary team of experts.

Professional inquiries are invited. East Coast residents, address Charles J. Fowler, Director of Admissions, Devereux Schools, Devon, Pennsylvania. West Coast residents, address Keith A. Seaton, Registrar, Devereux Schools, Box 1079, Santa Barbara, California. Southwestern residents, address John M. Barclay, Director of Development, Devereux Schools, Box 2269, Victoria, Texas.

THE DEVEREUX FOUNDATION

A nonprofit organization
Founded 1912
Devon, Pennsylvania
Santa Barbara, California
Victoria, Texas

SCHOOLS
COMMUNITIES
CAMPS
TRAINING
RESEARCH

HELENA T. DEVEREUX
Administrative Consultant

EDWARD L. FRENCH, Ph.D.
Director